Functional Genomics Studies of Atlantic Salmon (*Salmo salar*) Development, and Sac Fry Responses to Chronic Incremental Hyperthermia

By: Larina A. Carroll

A thesis submitted to the Department of Biology and

the School of Graduate Studies

in partial fulfilment of the requirements for the degree of

Master of Science

Ocean Sciences Centre and Department of Biology,

Memorial University of Newfoundland,

St. John's, Newfoundland and Labrador, Canada

ABSTRACT

Functional genomics techniques, including the 32K cGRASP microarray and real-time quantitative polymerase chain reaction (QPCR), were used to study gene expression in a single Atlantic salmon cohort: 1) during embryonic and early sac fry development; and 2) after sac fry were exposed to chronic incremental hyperthermia (water temperature increased by 1 °C every 24 hours, from 7.4 to 21.4 °C). The first study focused on the transcript expression of four microarray-identified paralogous *γM-crystallin* genes as well as four paralogous *α-sHSP* genes. The four *α-sHSPs,* as well as *cryGM4-like* and *cryGM3-like,* were shown by QPCR as higher expressed at hatch or post-hatch stage relative to eye-up stage (although the degree of induction varied between paralogues). This research provides possible evidence of divergent transcript expression (i.e., regulation) of duplicated genes, suggesting that some of the paralogues studied may have diverged functionally. These results provide insight into the evolutionary relationships between these genes, and may provide evidence of neofunctionalization following GD events. The second study focused on four microarray-identified genes of interest (*trypsin-1 precursor*, *chymotrypsin b*, *ferritin middle subunit*, and *ubiquitin*) as well as the four paralogous *α-sHSP* genes. The transcripts encoding *trypsin-1 precursor*, *ferritin middle subunit*, and *ubiquitin*, as well as *α-sHSPs 1*, *3* and *4*, were shown by QPCR as responsive to the incremental hyperthermia conditions. This study identified molecular biomarkers that may be useful for studying adaptation of early life stage salmon under potentially stressful conditions (e.g., warming climate).

ACKNOWLEDGEMENTS

I thank my supervisor, Dr. Matthew Rise, for his expertise, support, and encouragement throughout the course of this research. In addition, I would like to thank the staff of the Marine Institute's Aquaculture Facility, particularly Jason Nichols and Keith Rideout, who provided instruction in, and assistance with, gamete stripping and fertilization, and gave advice on aquaculture procedures during the rearing stages. Terry Bungay provided assistance in setting up the rearing tanks, advice during rearing, and technical training on the LaMotte SMART 2 Colorimeter (LaMotte Company, Chestertown, MD). Dr. Laura Halfyard also assisted with training on the colorimeter. Dr. Lourdes Peña-Castillo consulted on bioinformatics procedures. Dr. Marije Booman and Dr. Tiago Hori provided assistance and advice on functional genomics lab procedures. Charles Feng provided technical training and assisted with primer design, sequence alignment and molecular phylogenetics. Danny Ings and Charles Feng aided in, and advised on, statistical analysis of the QPCR data. Dr. Albert Caballero Solares advised on and conducted the statistical analysis of the water quality and mortality data in Chapter 3 of this thesis. I thank Dr. Harry Murray for his time, insight, and advice during the writing of this thesis. I acknowledge my supervisory committee, Dr. Matthew Rise, Dr. Paul Snelgrove, and Dr. Kurt Gamperl (Department of Ocean Sciences, MUN), who provided knowledgeable expertise, guidance, and feedback. This research was supported by an NSERC Discovery Grant to Dr. Matthew Rise.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

α – alpha

ANOVA – analysis of variance

ATF3 – activating transcription factor 3

BLAST – Basic Local Alignment Search Tool

BP – biological process

CC – cellular component

cDNA – complementary deoxyribonucleic acid

cGRASP – consortium for Genomic Research on All Salmonids Project

Contig – contiguous sequence

C_t – threshold cycle

Cy3 and Cy5 – Cyanine3 and Cyanine5

Db – database

Dd – degree days

DNA – deoxyribonucleic acid

DO – dissolved oxygen

dpf – days post-fertilization

EST – expressed sequence tag

E-value – expect value

ExPASy – Expert Protein Analysis System

FSGD or R3 – fish-specific (whole) genome duplication (event) (i.e. third round of genome duplication in the vertebrate lineage)

γ – gamma

GD or WGD – genome duplication or whole genome duplication

GO – gene ontology

GOI – gene of interest

HSP – heat shock protein

Indel – insertion/deletion

IPCC - The Intergovernmental Panel on Climate Change

MCH – melanin-concentrating hormone

MF – molecular function

MI – Marine Institute

MRF – myogenic regulatory factor

mRNA – messenger ribonucleic acid

MSA – multiple sequence alignment

M.y.a. – million years ago

NCBI – National Centre for Biotechnology Information

Nr – non-redundant

Nt – nucleotide

NTC – no-template control

ORF – open reading frame

OSC – Ocean Sciences Centre

PABP1 – Polyadenylate binding protein 1, or Poly-A binding protein 1

pIC – polyriboinosinic polyribocytidylic acid

PMT – photo-multiplier tube

QPCR – real-time quantitative polymerase chain reaction

R^2 – coefficient of determination

RACE – rapid amplification of cDNA ends

RNA – ribonucleic acid

RQ – relative quantification

SEM – standard error of the mean

sHSP – small heat shock protein

SNP – single nucleotide polymorphism

TAN – total ammonia nitrogen

TMS or MS-222 – tricaine methanesulfonate

UniProtKb – Universal Protein Resource Knowledgebase

UTR – untranslated region

CO-AUTHORSHIP STATEMENT

The research described in this thesis was conducted by Larina A. Carroll, under the supervision of Dr. Matthew L. Rise. Larina A. Carroll was responsible for rearing of the salmon, the described experiments, data collection and analysis. The Atlantic salmon *α-sHSP* paralogues were selected as genes of interest in this thesis based on a previous study which found that rainbow trout orthologues of these genes were more highly expressed surrounding the hatch event compared with earlier developmental stages (Rise and Devlin, unpublished). Manuscripts resulting from this thesis will be prepared by Larina A. Carroll, with editing assistance and intellectual input from co-authors as follows:

If Chapter 2 of this thesis leads to a publication, authorship on the manuscript would be Larina A. Carroll, Charles Y. Feng, Dr. Robert H. Devlin, and Dr. Matthew L. Rise.

If Chapter 3 of this thesis leads to a publication, authorship on the manuscript would be Larina A. Carroll, Dr. Albert Caballero-Solares, and Dr. Matthew L. Rise.

1. GENERAL INTRODUCTION

1.1 ATLANTIC SALMON (*SALMO SALAR*) LIFE HISTORY

Atlantic salmon, *Salmo salar,* spawn in the fall from late October through November, migrating up-river as they search for optimal spawning sites. Females use their tails to excavate shallow depressions, or redds, in the river bed substrate, in which they lay their eggs (Scott and Scott 1988). Once eggs have been laid, the male releases milt over the eggs to fertilize them, and the female again uses her tail to bury them in the substrate, where they will progress through various developmental stages (Scott and Scott 1988). The most readily identifiable of these stages is called eye-up, which occurs when the pigmentation of the eyes becomes visible through the chorion (Johnston *et al.* 1999). During the incubation period, the vitelline membrane, enveloping layer, and chorion protect the salmonid embryo from poor water quality and other environmental factors (Finn 2007). However, at hatch, the salmonid sac fry leave the protection of the egg and emerge into a physiologically demanding environment. Hatching around the end of April presents many types of environmental stressors which challenge immune, thermoregulatory, respiratory, and osmoregulatory functions (Finn 2007). At this point in their life cycle, the sac fry, or alevin [or eleuthero-embryo (Allen and Ritter 1977)], remain in the substrate until they develop further, obtaining nourishment by gradually reabsorbing the yolk sac (Scott and Scott 1988). Once the yolk sac has been absorbed, the salmon, now called fry, emerge from the substrate and enter the water column. As the salmon grow, they

develop into parr, recognized by lateral bar markings on their sides, and then into smolt (Scott and Scott 1988). At this point in their life cycle, most Atlantic salmon migrate to the ocean and remain there until they reach sexual maturity. However, some remain in freshwater rivers and streams.

1.2 FISH EARLY LIFE STAGE GENE EXPRESSION IN RESPONSE TO TEMPERATURE CHANGES

Wild Atlantic salmon face many challenges as they develop and grow from embryos to adults, and the early life history stages face harsh conditions and unique challenges associated with the breeding strategy of this species. Developmental duration is temperature dependent, and the time at which the Atlantic salmon sac fry emerge from the substrate varies with latitude, ranging from as early as May in warmer climates to as late as June in cooler climates (Scott and Scott 1988). The Intergovernmental Panel on Climate Change (IPCC 2007) predicts that, in addition to changes in ocean temperature and sea level, future winters in the Northern hemisphere will be milder and wetter, with hotter and drier summers and higher storm frequency. The effects of climate change will likely be exaggerated over land compared to over the oceans, resulting in a greater influence of this phenomenon on fresh water as opposed to marine systems (Bogner *et al.* 2008). As a result, Atlantic salmon, as well as other salmonids, will likely shift their population distributions toward the north, seeking cooler waters. Climate change may also alter the life history of salmonid species, accelerate development, increase mortality rates and

susceptibility to disease, and delay spawning as a result of longer, warmer summers (Jonsson and Jonsson 2009; Reist *et al.* 2006). Elevated temperature from climate change may also alter gene expression and physiology during early development (e.g., of embryo and sac fry stages). Given the importance of ambient temperature to the early development of Atlantic salmon (Albokhadaim *et al*. 2007; Ytteborg *et al.* 2010; Macqueen *et al.* 2008), understanding how elevated temperatures impact gene expression during early life-history stages could help to predict the ability of wild populations of this species to adapt to climate change.

Genomics techniques, such as microarray hybridizations, have been used to study the genes and molecular pathways involved in fish development (e.g., Evans *et al*. 2015; Jantzen *et al.* 2011), and responses to environmental and other stressors (e.g., Healy *et al.* 2010; Ikeguchi *et al.* 2006; Hori *et al.* 2012). Previous work has examined the response of fishes to temperature change in terms of body size and muscle cellularity. This research showed that salmon embryos or sac fry reared at higher temperatures exhibit higher growth rates than counterparts reared at comparatively lower temperatures. This was despite the fact that the elevated growth period was brief, and resulted in less dense muscle tissues (Albokhadaim *et al*. 2007; Ytteborg *et al.* 2010). In addition, this research showed that optimal muscle growth was observed in adult Atlantic salmon incubated at 5 °C as embryos (Macqueen *et al.* 2008). Northern Blot analysis has been used to show that temperature appears to influence the transcript expression of myogenic regulatory factors (e.g., MyoD and

myogenin) at hatching and yolk sac resorption stages in rainbow trout (*Oncorhynchus mykiss*) and sea bass (*Dicentrarchus labrax*) (Wilkes *et al.* 2001). Macqueen *et al.* (2007) also used *in situ* hybridization to show that the transcript expression of myogenic regulatory factors (e.g., MyoD family members) in Atlantic salmon somite formation was influenced by different incubation temperatures, with some transcripts exhibiting different temporal expression patterns. Thus, a size advantage, and decreased predation risk, may be conferred on young salmon reared at higher temperatures (Macqueen *et al.* 2007, 2008; Albokhadaim *et al*. 2007; Ytteborg *et al.* 2010). However, prior to this thesis research, little was known regarding Atlantic salmon gene expression surrounding the hatch event, or the effect of elevated ambient temperature on the transcriptome [i.e., the suite of all RNA molecules that determine the phenotype of an organism (Velculescu *et al.* 1997)]. In this thesis, microarrays were used to identify paralogues (i.e., related genes that arise by duplication of an ancestral gene) that were developmentally regulated (i.e., more highly expressed at hatch compared with eye-up stage) in order to examine the potential importance of gene duplication and divergence in this species. Additionally, a suite of Atlantic salmon *α-sHSP* paralogues were examined to determine if these defence-related genes are synchronously expressed surrounding the hatch event.

1.3 WHOLE GENOME DUPLICATION EVENTS

The Atlantic salmon diverged from its closest relative with a sequenced genome, the zebrafish (*Danio rerio*), about 200 million years ago (m.y.a.) (Jaillon *et*

al. 2004), following the whole genome duplication (WGD) event during early teleost evolution (Rose *et al.* 2008; Steinke *et al.* 2006). Fish genomes have undergone rediploidization since the fish-specific WGD, reducing the redundancy of paralogues. However, a salmonid-specific WGD event occurred approximately 80 m.y.a., and subsequent rediploidization has resulted in a pseudotetraploid salmonid genome [i.e., the diploid state of the genome has not been fully reestablished following whole genome duplication event(s)] (Allendorf and Thorgaard 1984; Li *et al.* 2007; Lien *et al*. 2016; Ravi and Venkatesh 2008). Genome duplication (GD) and rediploidization are believed to generate genetic diversity, leading to the species-specific functionalization of genes (Jaillon *et al.* 2004). The presence of a gene duplicate can alter evolutionary constraints on the ancestral gene; some pairs of paralogues diverge faster than others (Leong *et al.* 2010), and divergence of gene duplicates may be at the structural (sequence) level, at the gene or protein expression level, and/or at the functional level. The pseudotetraploid salmonid genomes are, therefore, of particular interest in the field of molecular evolution. This thesis presents data on the structure and expression of two sets of developmentally regulated paralogues in order to provide insight into gene or genome duplication and divergence, where new members of an expanded gene family (i.e., new paralogues) may: a) retain the expression profile of the ancestral gene, b) acquire new expression characteristics and functions (e.g., neofunctionalization), or c) become silenced (i.e., become a pseudogene) (Rose *et al*. 2008).

1.4 FUNCTIONAL GENOMICS TOOLS

The development of microarray technology has enabled high-throughput detection of differentially expressed genes between conditions, treatments, or life stages. For example, expression profiling of Atlantic salmon embryogenesis using the Agilent 44K oligonucleotide salmonid microarray platform produced a gene set of thousands of differentially expressed genes, with functional annotations such as blastocyst, brain, organ and erythrocyte development, as well as induction of apoptosis (Jantzen *et al.* 2011). The functional genomics studies presented in this thesis used the consortium for Genomic Research on All Salmonids Project (cGRASP) 32K (salmonid) cDNA microarray (Koop *et al.* 2008), which represented the best coverage of the salmonid transcriptome available at the time of the microarray-based component of this research. The availability of high-throughput tools such as microarrays makes the development and use of appropriate bioinformatics tools essential for the interpretation of the large data sets obtained in these experiments. Bioinformatics techniques can enable identification of functional annotations of genes of interest, the analysis of microarray data, multiple sequence alignment and phylogenetic tree construction to determine evolutionary relatedness, domain identification, and protein structure prediction (Booman and Rise 2012). In the research presented in this thesis, bioinformatics techniques were used to obtain functional annotations of transcripts, retrieve, compile, and characterize the sequence structure of paralogues *in silico*, as well as to determine similarity and evolutionary

relationships among paralogous sequences. Real-time quantitative polymerase chain reaction (QPCR), another functional genomics technique, can be used to determine biological variability of gene expression between individuals and over time (e.g., Feng and Rise, 2011; Rise *et al.* 2004, 2008), and is also often used to validate microarray results (Booman *et al.* 2011; Hori *et al.* 2012). In this thesis, QPCR was used to provide expression information for selected Atlantic salmon genes of interest (GOI).

1.5 RESEARCH OBJECTIVES

The main goal of this research was to utilize bioinformatics, molecular biology, and functional genomics tools and techniques to conduct two studies. The first study (Chapter 2 of this thesis) used the 32K microarray and QPCR to study the expression of sets of paralogues to determine if there was potential evidence of gene duplication and divergence (i.e., differences in transcript expression of paralogues) during early development in Atlantic salmon. Microarray hybridization identified a set of γ-crystallin-domain containing paralogues whose expression was higher at hatch relative to the eye-up stage. A previous study (Rise and Devlin, unpublished) indicated that four paralogous *α-sHSP* transcripts were more highly expressed just prior to the hatch event in rainbow trout embryos. These genes were functionally annotated as stress-responsive, and the researchers hypothesized that this pattern was a pre-adaptation to hatching stress when the embryos leave the protection of the egg and emerge as sac fry. Therefore, an additional objective of the first study was to

determine if *α-sHSP* paralogues were also more highly expressed in Atlantic salmon embryos surrounding the hatch event relative to eye-up (i.e., to examine if this developmental profile of expression is conserved within Salmonidae). This work provides insight into the expression profiles and structural divergence of selected members of highly radiated gene families. Although the salmon genome has been studied and sequenced (Davidson *et al*. 2010; Koop *et al.* 2008; Lien *et al.* 2016), little was known about global gene expression surrounding the salmonid hatch event prior to this research (Jantzen *et al.* 2011; Rise *et al.* 2007). Information acquired throughout this research provides a broader understanding of the genes, proteins, and molecular pathways that may be involved in salmonid hatching.

The objectives of the second study (Chapter 3 of this thesis) were to use the 32K microarray and QPCR to examine defence-relevant gene expression in Atlantic salmon sac fry in response to chronic incremental hyperthermia, and to identify molecular biomarkers involved in coping with hyperthermia at this developmental stage. In addition, the expression profiles of four Atlantic salmon *α-sHSP* paralogues were studied to determine if these transcripts are thermally responsive. Wild populations of salmon could experience fluctuations in ambient temperature associated with spring hatching, as well as elevated temperature due to climate change and other anthropogenic influences on their habitat. The QPCR assays developed in this research may be valuable in future assessments of the impact of temperature stress on early life stage farmed and wild salmonids.

1.6 REFERENCES

Albokhadaim, I., Hammond, C. L., Ashton, C., Simbi, B. H., Bayol, S., Farrington, S., Stickland, N. (2007). Larval programming of post-hatch muscle growth and activity in Atlantic salmon (*Salmo salar*). *Journal of Experimental Biology* 210: 1735-1741.

Allen, K. R., and J. A. Ritter (1977). Salmonid terminology. *Journal du Conseil / Conseil Permanent International pour l'Exploration de la Mer* 37(30):293-299.

Allendorf, F. W. and Thorgaard, G.H. (1984). Tetraploidy and the evolution of salmonid fishes. In: Evolutionary Genetics of Fishes (ed. B.J. Turner), Plenum Press, New York, 1-53.

Anttila, K., Eliason, E. J., Kaukinen, K. H., Miller, K. M., Farrell, A. P. (2014). Facing warm temperatures during migration: cardiac mRNA responses of two adult *Oncorhynchus nerka* populations to warming and swimming challenges. *Journal of Fish Biology* 84(5):1439-1456.

Bogner, J., Pipatti, R., Hashimoto, S., Diaz, C., Mareckova, K., Diaz, L., *et al.* (2008). Mitigation of global greenhouse gas emissions from waste: conclusions and strategies from the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report. Working Group III (Mitigation). *Waste Management and Research*, 1:11-32.

Booman, M., Borza, T., Feng, C.Y., Hori, T.S., Higgins, B., Culf, A., *et al.* (2011). Development and experimental validation of a 20K Atlantic cod (Gadus morhua) oligonucleotide microarray based on a collection of over 150,000 ESTs. *Marine Biotechnology* 13(4):733-750.

Booman, M., Rise, M.L. (2012). Genomic Tools for Understanding the Molecular Basis of Production-Relevant Traits in Finfish. In: Aquaculture Biotechnology (Eds. Fletcher, G.L., Rise, M.L.), Wiley-Blackwell, Oxford, UK, 3-19.

Cooke, S. J., Hinch, S. G., Donaldson, M. R., Clarke, T. D., Eliason, E. J., Crossin, G. T., *et al.* (2012). Conservation physiology in practice: how physiological knowledge has improved our ability to sustainably manage Pacific salmon during up-river migration. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 367(1596):1757-1769.

Davidson W.S., Koop B.F., Jones S.J., Iturra P., Vidal R., Maass A., *et al.* (2010). Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biology* 11:403.

Evans, M.L., Hori, T.S., Rise, M.L., Flemming, I.A. (2015). Transcriptomic responses of Atlantic salmon (*Salmo salar*) to environmental enrichment during juvenile rearing. *PLoS One* 10(3):e0118378.

Feng, C.Y., Rise, M.L. (2011). Identification and molecular cloning of Atlantic cod (Gadus morhua) activating transcription factor 3 (ATF3) transcript and its

induction in spleen following intraperitoneal polyriboinosinic polyribocytidylic acid injection. *Fish and Shellfish Immunology* 31(3): 475-481.

Finn, R. N. (2007). The physiology and toxicology of salmonid eggs and larvae in relation to water quality criteria. *Aquatic Toxicology* 8(4): 337-354.

Healy, T.M., Tymchuk, W.E., Osborne, E.J., Schulte, P.M. (2010). Heat shock response of killifish (*Fundulus heteroclitus*) candidate gene and heterologous microarray approaches. *Physiological Genomics* 41:171-184.

Heredia-Middleton, P., Brunelli, J., Drew, R.E., Thorgaard, G.H. (2008). Heat shock protein (HSP70) RNA expression differs among rainbow trout (*Oncorhynchus mykiss*) clonal lines. *Comparative Biochemistry and Physiology, Part B: Biochemistry and Molecular Biology* 149: 552-556.

Hori, T.S., Gamperl, A.K., Booman, M., Nash, G.W., Rise, M.L. (2012). A moderate increase in ambient temperature modulates the Atlantic cod (*Gadus morhua*) spleen transcriptome response to intraperitoneal viral mimic injection. *BMC Genomics* 13(1):431.

Ikeguchi, K., Ineno, T., Itoi, S., Kondo, H., Kinoshita, S., Watabe, S. (2006). Increased levels of mitochondrial gene transcripts in the thermally selected rainbow trout (*Oncorhynchus mykiss*) strain during embryonic development. *Marine Biotechnology* 8: 178-188.

Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N., Mauceli, E., *et al.* (2004). Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* 431: 946-957.

Jantzen, S.J., Sanderson, D.S., von Schalburg, K.R., Yasuike, Marass, F., Koop, B.F. (2011). A 44K microarray dataset of the changing transcriptome in developing Atlantic salmon (*Salmo salar* L.) *BMC Research Notes* 4:88.

John, LB, Yoong, S, Ward, AC. (2009). Evolution of the Ikaros gene family: implications for the origins of adaptive immunity. *Journal of Immunology* 182: 4792-4799.

Johnston, I.A., Strugnell, G., McCracken, M.L., Johnstone, R. (1999). Muscle growth and development in normal-sex-ratio and all-female diploid and triploid Atlantic salmon. *Journal of Experimental Biology,* 202(15): 1991-2016.

Jonsson, B., Jonsson, N. (2009). A review of the likely effects of climate change on anadromous Atlantic salmon *Salmo salar* and brown trout *Salmo trutta*, with particular reference to water temperature and flow. *Journal of Fish Biology* 75:2381-2447.

Koop, B.F., von Schalburg, K.R., Leong, J., Walker, N., Lieph, R., Cooper, G.A., *et al.* (2008). A salmonid EST genomic study: genes, duplications, phylogeny and microarrays. *BMC Genomics* 9: 545.

Lang, L., Miskovic, D., Fernando, P., Heikkila, J.J. (1999). Spatial pattern of constitutive and heat shock-induced expression of the small heat shock protein gene family, Hsp30, in *Xenopus laevis* tailbud embryos. *Developmental Genetics* 25:365-374.

Leong, J.S., Jantzen, S.G., von Schalburg, K.R., Cooper, G.A., Messmer, A.M., Liao, N.Y., *et al.* (2010). *Salmo salar* and *Esox lucius* full-length cDNA sequences reveal changes in evolutionary pressures on a post-tetraploidization genome. *BMC Genomics* 11:279.

Li, C., Ortí, G., Zhang, G., Lu, G. (2007). A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evolutionary Biology* 7:44.

Lien, S., Koop, B.F., Sandve, S.R., Miller, J.R., Kent, M.P., Nome, T., *et al.* (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature* 533:200-205.

Macqueen, D.J., Robb, D., Johnston, I.A. (2007). Temperature influences the coordinated expression of myogenic regulatory factors during embryonic myogenesis in Atlantic salmon (*Salmo salar* L.). *The Journal of Experimental Biology* 210:2781-2794.

Macqueen, D.J., Robb, D., Olsen, T., Melstvelt, L., Paxton, C., Johnston, I.A. (2008). Temperature until the 'eyed stage' of embryogenesis programs the

growth trajectory and muscle phenotype of adult Atlantic salmon. *Biology Letters 4:294-298.*

Metz, B., Davidson, O.R., Bosch, P.R., Dave, R., Meyer, L.A. (Eds.) (2007). Contribution of working group III to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom and New York, USA.

Mosser, D.D., Bols, N.C. (1988). Relationship between heat-shock protein synthesis and thermotolerance in rainbow trout fibroblasts. *Journal of Comparative Physiology* 148: 457-467.

Ohan, N.W., Heikkila, J.J. (1995). Involvement of differential gene expression and mRNA stability in the developmental regulation of the hsp 30 gene family in heat-shocked *Xenopus laevis* embryos. *Developmental Genetics* 17:176.

Pérez-Casanova, J.C., Rise, M.L., Dixon, B., Afonso, L.O., Hall, J.R., Johnson, S.C., Gamperl, A.K. (2008). The immune and stress responses of Atlantic cod to long-term increases in water temperature. *Fish and Shellfish Immunology* 24:600-609.

Purcell, M.K., Laing, K.J., Woodson, J.C., Thorgaard, G.H., Hansen, J.D. (2009). Characterization of the interferon genes in homozygous rainbow trout reveals two novel genes, alternate splicing and differential regulation of duplicated genes. *Fish and Shellfish Immunology* 26:293-304.

Purcell, M.K., Nichols, K.M., Winton, J.R., Kurath, G., Thorgaard, G.H., Wheeler, P., *et al.* (2006). Comprehensive gene expression profiling following DNA vaccination of rainbow trout against infectious hematopoietic necrosis virus. *Molecular Immunology* 43:2089-2106.

Ravi, V., Venkatesh, B. (2008). Rapidly evolving fish genomes and teleost diversity. *Current Opinion in Genetics & Development* 18: 544-550.

Reist, J.D., Wrona, F.J., Prowse, T.D., Power, M., Dempson, J.B., King, J.R., Beamish, R.J. (2006). An overview of effects of climate change on selected freshwater and anadromous fishes. *AMBIO: A Journal of the Human Environment* 35(7); 381-387.

Rise, M.L., Hall, J., Rise, M., Hori, T., Gamperl, A.K., Kimball, J., *et al.* (2008). Functional genomic analysis of the response of Atlantic cod (*Gadus morhua*) spleen to the viral mimic polyriboinosinic polyribocytidylic acid (pIC). *Developmental ad Comparative Immunology* 32(8): 916-931.

Rise, M.L., Jones, S.R.M., Brown, G.D., von Schalburg, K.R., Davidson, W.S., Koop, B.F. (2004). Microarray analyses identify molecular biomarkers of Atlantic salmon macrophage and hematopoietic kidney response to *Piscirickettsia salmonis* infection. *Physiological Genomics* 20:21-35.

Rise, M. L., Liu, Z., Douglas, S. E., Brown, L. L., Nash, J. H. E., McFall-Ngai, M. J. (2009). Aquaculture-related applications of DNA microarray technology. In: Molecular Research in Aquaculture (Ed. K. Overturf), Wiley-Blackwell,

Oxford, UK, 65-101.

Rise, M. L., von Schalburg, K. R., Cooper, G. A., Koop, B. F. (2007). Salmonid DNA microarrays and other tools for functional genomic research. In: Aquaculture Genome Technologies (Ed. Z. Liu), Blackwell Publishing Ltd., Oxford, UK, 369-411.

Rose, D., Jöris, J., Hackermüller, J., Reiche, K., Li, Q., Stadler, PF. (2008). Duplicated RNA genes in teleost fish genomes. *Journal of Bioinformatics and Computational Biology* 6: 1157-1175.

Scott, W.B., Scott, M.G. (1988). Atlantic Fishes of Canada. *Canadian Bulletin of Fisheries and Aquatic Sciences* 219:129-134.

Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E. Jr., *et al.* (1997). Characterization of the yeast transcriptome. *Cell* 88(2):243-51.

von Schalburg, K.R., Rise, M.L., Cooper, G.A., Brown, G.D., Gibbs, A.R., Nelson, C.C., *et al.* (2005). Fish and chips: various methodologies demonstrate utility of a 16,006-gene salmonid microarray. *BMC Genomics* 6:126.

Voyer, J., Heikkila, J.J. (2008) Comparison of the effect of heat shock factor inhibitor, KNK437, on heat shock- and chemical stress-induced hsp30 gene expression in *Xenopus laevis* A6 cells. *Comparative Biochemistry and Physiology, Part A: Molecular and Integrative Physiology* 151A:253-261

Vuori, K.A., Koskinen, H., Krasnov, A., Koivumäki, P., Afanasyev, S., Vuorinen, P.J.,

Nikinmaa, M. (2006). Developmental disturbances in early life stage mortality (M74) of Baltic salmon fry as studied by changes in gene expression. *BMC Genomics* 7: 56.

Wilkes, D., Xie, S. Q., Stickland, N. C., Alami-Durante, H., Kentouri, M., Sterioti, A., *et al.* (2001). Temperature and myogenic factor transcript levels during early development determines muscle growth potential in rainbow trout *(Oncorhynchus mykiss)* and sea bass *(Dicentrarchus labrax). The Journal of Experimental Biology* 204: 2763-2771.

Xie, Y., Song, L., Weng, Z., Liu, S., Liu, Z. (2015). Hsp90, Hsp60, and sHSP families of heat shock protein genes in channel catfish and their expression after bacterial infections. *Fish and Shellfish Immunology* 44(2): 642-651.

Ytteborg, E., Baeverfjord, G., Torgersen, J., Hjelde, K., Takle, H. (2010). Molecular pathology of vertebral deformities in hyperthermic Atlantic salmon (*Salmo salar*). *BMC Physiology* 10:12.

2. FUNCTIONAL GENOMICS STUDY OF ATLANTIC SALMON DEVELOPMENT SURROUNDING THE HATCH EVENT

2.1 ABSTRACT

The 32K cGRASP microarray platform was used to identify 73 transcripts that were reproducibly more highly expressed in hatch stage embryos compared to eye-up stage embryos, including features representing four γM-crystallin-domain-containing paralogues. QPCR assays were then developed and used to study the mRNA expression of the four *γM-crystallin* genes as well as four paralogous *α-sHSP* genes at developmental stages from eye-up to post-hatch. For both the *γM-crystallin* and the *α-sHSP* sequences, structural similarity (percent identity at the predicted amino acid level) was highest for pairs of paralogues that had similar transcript expression profiles (e.g., *α-sHSPs 1* and *2* transcript expression peaked at 57 dpf while peak expression of *α-sHSP 3* and *α-sHSP 4* occurred at 61 dpf). This study provides evidence of potential neofunctionalization of duplicated genes and insights on the ontogeny of gene expression during late embryogenesis and hatch in Atlantic salmon.

2.2 INTRODUCTION

Paralogues are members of gene families that arise as a result of gene or whole genome duplication (WGD) events and may either retain the attributes of the ancestral gene, acquire novel attributes, or become silenced (Bailey *et al.* 1978;

Sidow 1996; Ohno 1970). Atlantic salmon *(Salmo salar*) have experienced two WGD events over the past ~350 million years (Davidson *et al.* 2010; Steinke *et al.* 2006a, b). The first, a fish-specific WGD (FSGD, or R3; the third round of WGD in the vertebrate lineage) (Myer and Schartl 1999), occurred about 320-350 million years ago (m.y.a.) in the teleost lineage and gave rise to extant Actinopterygian (ray-finned) fishes, which have since undergone rediploidization (Allendorf and Thorgaard 1984; Christoffels *et al.* 2004; Meyer and Van de Peer 2005; Ohno 1970; Taylor *et al.* 2003). The second, more recent WGD event occurred around 80 m.y.a. in the order Salmoniformes and gave rise to extant salmonids including salmon and trout species, which are considered pseudotetraploid (Allendorf and Thorgaard 1984; Lien *et al.* 2016). WGD and rediploidization are thought to generate genetic diversity and to lead to species-specific neofunctionalization of genes, such that some paralogues may acquire different functions from those of their ancestral sequences (Ohno 1970; Ohno *et al.*1968; Wolfe 2001). Studying the transcriptome of a pseudotetraploid species such as Atlantic salmon may provide insights into gene duplication and divergence, the WGD, and the potential role of these events in adaptation, speciation, and evolution.

In this study, microarray analysis was used to compare global transcript expression, and to identify defence-relevant genes and suites of putative paralogues that are differentially expressed between eye-up and hatch stage embryos. Further, the current study utilizes QPCR to examine the expression profiles of a microarray-

identified suite of paralogues (*γM-crystallin-like* transcripts) and four paralogous *α-sHSP* transcripts during late embryogenesis and surrounding the hatch event in Atlantic salmon. The *γM-crystallin-like* and *α-sHSP* paralogues were selected for developmental QPCR analysis to look for evidence of gene duplication and divergence (i.e., different gene expression profiles of paralogues) because there was evidence of multiple paralogues and differential developmental regulation (i.e., from the microarray experiment for *γM-crystallin-like* paralogues, and from Rise and Devlin unpublished for the *α-sHSP* paralogues). In this study, bioinformatic techniques were used to compare the structure (nucleotide and hypothetical amino acid sequence) and transcript expression of *γM-crystallin-like* and *α-sHSP* paralogues within Atlantic salmon. This research builds upon publicly available sequence databases, as well as functional genomic resources (e.g., the 32K microarray) for Atlantic salmon (Rise *et al.* 2004b; von Schalburg *et al.* 2005, 2008a; Koop *et al*. 2008), a commercially and ecologically important species. This research also addresses whether these suites of paralogues retain common (potentially ancestral) regulation or if they have adopted new patterns of regulation (potentially suggesting neofunctionalization), and improves our understanding of the evolution of these salmonid gene families.

2.3 METHODS

2.3.1 Gamete stripping and fertilization

Adult Atlantic salmon from a broodstock population held in outdoor tanks (4,000 L, fresh water, ~7 °C) at the Marine Institute (MI) of Memorial University of Newfoundland were stripped for eggs and milt on November 26th, 2008. Males and females were individually captured and placed in an anaesthetic bath [75 mg•L^{-1} of tricaine methanesulfonate (TMS, or MS-222, Syndel Laboratories Ltd., B.C. Canada)] buffered with sodium bicarbonate (75 mg•L^{-1}, to attain a neutral pH). The fish were stripped and placed in a recovery bath for observation before they were returned to their holding tanks. All implements and tools were sterilized with iodine and rinsed with distilled water before use, and between individual fish.

Following stripping, the milt was examined for motility by activating a sample with a drop of ~7 °C MI well water followed by observation under a compound microscope. The cohort used in this experiment was obtained by fertilizing eggs (about 1 L) from a single female with motile milt from two males (about 25 mL from each male). The fertilized eggs were treated with 1% Ovadine (Western Chemicals, Ferndale, WA; a general fish egg surface disinfectant) diluted in ~7 °C MI well water for ten min to remove any ectoparasites or pathogens, rinsed thoroughly in MI well water, and divided equally among three baskets in monolayers. These baskets were constructed in the following manner to hold the

embryos: 1) PVC tubing (20 cm diameter) was cut in 10 cm sections; 2) mesh (2 mm netting) was cut to fit around one end of the tubing, and secured in place using rubber bands to create a mesh-bottomed basket; and 3) plastic legs were attached to the tubes to elevate the mesh about 5 cm above the bottom of the rearing tank. Triplicate tanks were set up with one basket in each tank. Each triplicate tank was equipped with an air stone, and received 7 °C MI well water that entered from underneath the baskets. Embryos and sac fry were kept in the dark at all times, except during sampling.

2.3.2 Water quality

The flow-through water in the tanks was obtained from the MI well. Tank conditions were monitored every 24 h by measuring the water temperature (using a digital thermometer), percent dissolved oxygen saturation (DO) (using a DO probe and meter; OxyGuard Point Four Systems Inc., BC, Canada), and total ammonia nitrogen (TAN) [using a LaMotte SMART 2 Colorimeter (LaMotte Company, Chestertown, MD)]. Un-ionized ammonia (NH_3 in $mg \bullet L^{-1}$) levels were calculated using Equation 1 (Web Reference 2.10 and 2.12).

Equation 1: $NH_3 = (TAN)\ (f)\ (\text{LaMotte constant})$

The LaMotte constant is 1.216. This constant [obtained by dividing the molar mass of ammonia (17.031 $g \bullet mol^{-1}$) by the molar mass of ammonia as nitrogen (14.007 $g \bullet mol^{-1}$)] converts the result from un-ionized ammonia-nitrogen to un-ionized

ammonia (NH_3) (Web Reference 2.12). In this equation, *f* is the fraction of total ammonia that is un-ionized, and is calculated using Equation 2 (as described in Emerson *et al.* 1975; Web Reference 2.11 and 2.12).

Equation 2: $f = 1 / [10^{(pK-pH)} + 1]$

The pH of the freshwater obtained from the MI well was 7.5 for the duration of the experiment. In Equation 2, pK refers to the equilibrium constant between un-ionized and ionized ammonia and is calculated using Equation 3 (as described in Emerson *et al.* 1975; Web Reference 2.11 and 2.12).

Equation 3: pK = 0.09018 + 2729.2 / (temperature °C + 273.2)

One-way repeated measures analysis of the variance (ANOVA) were conducted to determine any time points with statistically significant ($p < 0.05$) differences in water quality.

2.3.3 Sampling

Every 24 h, mortalities were removed from each of the replicate tank baskets. Every 48 h, two embryos or sac fry were sampled from each triplicate tank (although only selected time points were used in this study), placed in individual nuclease-free microcentrifuge tubes, flash-frozen in crushed dry ice, and transported from the MI to the Rise Lab (Ocean Sciences Centre, St. John's, NL) where they were stored at -80 °C until use. All utensils and tools used for sampling were disinfected with iodine and rinsed with distilled water, followed by RNase Away (Molecular

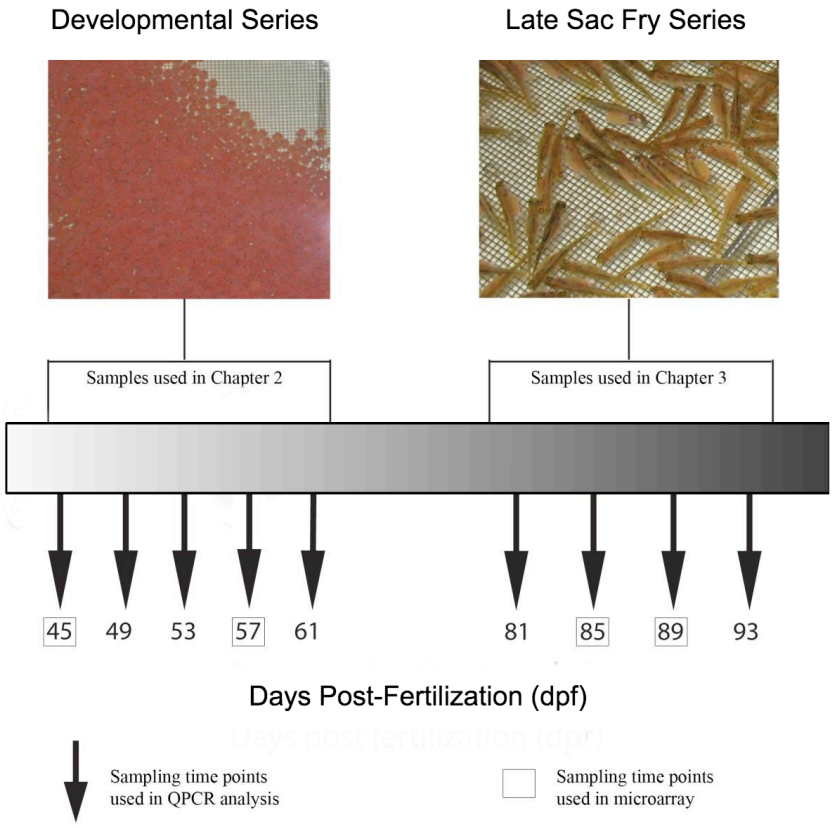
BioProducts, San Diego, CA), before use and between individuals. The sampling time points selected for gene expression analyses were: 1) the eye-up stage at 45 days post-fertilization (dpf) [333 degree days (dd): dd = (dpf)(temperature in °C)], using the average water temperature of 7.45 °C); 2) the fin-ray development stage at 49 dpf (363 dd) (as described by Johnston *et al*. 1999); 3) pre-hatch (53 dpf; 392 dd); 4) 50% hatch (57 dpf; 422 dd); and 5) post-hatch (61 dpf; 451 dd) (see Figure 2.1).

2.3.4 RNA extraction and purification

Individual embryo or sac fry were ground to a fine powder under liquid nitrogen using mortars and pestles. Before use, and between individual RNA extractions, ceramic mortars and pestles were cleaned by washing with a dilute soap solution and rinsing with water, followed by soaking overnight in 10% bleach. Mortars and pestles were rinsed thoroughly and repeatedly in MilliQ water, inverted, and left to dry overnight in a fume hood. Finally, mortars and pestles were wrapped in aluminium foil and baked at 220 °C for five h to inactivate RNases.

To extract the RNA from the homogenized tissue, 800 µl of TRIzol reagent (Invitrogen, Carlsbad, CA) was added to the ceramic mortar and mixed with the frozen tissue powder, each sample was passed through a QIAshredder (QIAGEN, Mississauga, ON) following the manufacturer's instructions, and the manufacturer's protocol was followed for TRIzol-based RNA isolation. Finally, each RNA pellet (representing RNA from an individual embryo or sac fry) was re-suspended in 12 µl

Figure 2.1 Experimental overview, showing sampling time points in days post-fertilization (dpf). Eye-up occurred at approximately 45 dpf, and 50% hatch occurred at approximately 57 dpf. Four late sac fry developmental time points were also sampled for a separate study (Chapter 3 of this thesis). Late sac fry still possessed yolk sacs, and did not receive exogenous feeding, although swim-up was expected to occur shortly after the conclusion of the studies.


Developmental Series
Late Sac Fry Series
Samples used in Chapter 2
Samples used in Chapter 3
45
49
53
57
61
81
85
89
93
Days Post-Fertilization (dpf)
Sampling time points used in QPCR analysis
Sampling time points used in microarray

of nuclease-free water (Invitrogen) and placed in a -80 °C freezer.

Using the manufacturer's instructions, total RNA (30 µg from each individual) was: 1) digested with 6.8 Kunitz units of DNase 1 (RNase-Free DNase Set, QIAGEN) to degrade genomic DNA; 2) column-purified using the RNeasy MinElute Cleanup Kit (QIAGEN) to remove traces of DNase 1, salt, or other impurities; and 3) eluted with 20 µl of nuclease-free water. Purified RNA samples were kept at -80 °C until use.

Total RNA quantity, purity and integrity were assessed for both pre-cleaned and post cleaned samples using A260/280 and A260/230 NanoDrop ND1000 UV spectrophotometry (with 1 µl used per individual RNA sample, and accepted ratios of approximately 1.8-2.0 for A260/280 and 2.0-2.2 for A260/230), and 1% agarose gel electrophoresis. Gels were stained with ethidium bromide and run in 1x TAE buffer; 1 µg of each individual RNA sample was electrophoretically separated alongside a 1 Kb Plus DNA ladder (Invitrogen). These gels were visualized with a UV transilluminator in a G:BOX (Syngene, Frederick, ML).

2.3.5 Microarray hybridization and data acquisition

Direct comparison microarray experiments using pooled RNA templates (similar to the experimental design used in this thesis) have previously been shown to be effective in identifying differentially expressed transcripts for subsequent QPCR studies (e.g., Hall *et al.* 2011, Rise *et al.* 2004a, Rise *et al.* 2006). Therefore, a

direct comparison microarray experiment was run using the cGRASP 32K salmonid cDNA microarray (Koop *et al*. 2008). This experiment used three technical replicate microarray slides, including one dye-swap (to account for dye bias), and pooled RNA templates to identify putative paralogous transcripts differentially expressed between the eye-up stage (45 dpf) and 50% hatch (57 dpf) embryos for the QPCR study (see Figure 2.3). Sub-samples of purified RNA from the individual with the highest quality and yield from tank one and tank two at 45 dpf or 57 dpf were combined in 6 µg pools (3 µg per individual) and used to generate fluorescently labeled targets for the microarray experiment. Individuals from tank three did not contribute to the pooled RNA samples used in the microarray, but were included in the QPCR experiment (i.e., in order to add additional, previously un-examined, biological replicates, allowing for the discovery of novel information at the quantitive stage). In this experimental design, the microarray served as a tool for identifying dysregulated genes for QPCR analysis. Additionally, this conservative microarray design was necessitated by restrictions and difficulties caused by spatial effects, as described below.

The Array 900 Detection Kit (Genisphere, Hatfield, PA) (Cy3 and Cy5) and SuperScript II (Invitrogen) were used for microarray target synthesis and subsequent hybridization, respectively, following the manufacturer's instructions (as in Hall *et al.* 2011), with 1 µg of pooled total RNA used for the synthesis of each target. Microarrays were prepared for hybridization by washing twice in 0.1% SDS (10 min

per wash), followed by 5 x 1 min washes in MilliQ water. In an attempt to prevent "black hole" spatial effects (i.e., areas on the microarray where features have little or no fluorescent signal due to high background) (Bowtell and Sambrook 2003), arrays were also incubated in a hybridization oven at 50 °C for 45 min in a blocking solution of 5x SSC/0.1% SDS/0.2% BSA, followed by two washes (20 sec., at room temperature) in MilliQ water. The arrays were dried by centrifugation (514 x *g* for five min at room temperature) and placed in the hybridization oven (50 °C) until needed. Hybridization to the array occurred under LifterSlips (Thermo Scientific, Waltham, MA) which were washed in mild hand soap, rinsed in nuclease-free water, 70% ethanol, and 100% ethanol (molecular biology grade; Commercial Alcohols, Brampton, ON), dried using Kimwipes (Kimberly-Clark, Mississauga, ON), and finally sprayed with canned compressed air (Fisher Scientific) to remove dust from the LifterSlip's surface. Hybridizations were run in the dark in hybridization chambers (Corning, Corning, NY, USA) placed in a 50 °C water bath in the hybridization oven. Following hybridization (approximately 16 h), the LifterSlips were 'floated' off the array in pre-warmed (50 °C) 2x SSC/0.2% SDS. Arrays were washed once for 15 min at 50°C in 2x SSC/0.2% SDS, once for 15 min at room temperature in 2x SSC, once for 15 min at room temperature in 0.2x SSC, and once for 5 min at room temperature in 2x SSC. All wash buffers were made using nuclease-free water (Life Technologies). Arrays were dried by centrifugation (514 x g, 5 min, at room temperature) in 50 mL conical tubes. Arrays were scanned

immediately to obtain fluorescent images using ScanArray Express (PerkinElmer, Waltham, MA). Microarrays were scanned at 10 μm resolution with laser power set to 90%; the same photo-multiplier tube (PMT) settings were used for all slides in the study (PMT 75 Cy3, PMT 70 Cy5). Although additional arrays were hybridized and scanned, only three microarrays were free of "black hole" spatial effects and were therefore utilized for data extraction and analysis.

2.3.6 Microarray data analysis

ImaGene v7.5 software (Biodiscovery, El Segundo, CA) was used to extract signal intensity data from TIFF images corresponding to each channel of each scanned microarray. In order to increase confidence in the placement of the grid during data extraction in ImaGene v7.5, grids were manually placed over the TIFF images of each array, spots adjusted automatically, spatial effects were flagged manually, and the data extracted and analyzed three separate times. All three gene lists for a given comparison (i.e., > 2-fold lower expressed in 50% hatch compared with eye-up stage and > 2-fold higher expressed in 50% hatch compared with eye-up stage) obtained by separate grid placement, data extraction, and analysis were consistent, and the gene lists were therefore accepted with a high degree of confidence despite the presence of some spatial effects (e.g., areas of higher background signal on the arrays). To extract the data, raw image data for each channel (Cy3 and Cy5) of the array was uploaded into ImaGene 7.5. Transcript identification (ID) information was assigned to each spot in the array using a gene ID

file and a detailed annotation file (publicly available through cGRASP, see Web Reference 2.3). As in Hall *et al.* (2011), extracted microarray data were analyzed in the following manner using GeneSpring v7.3 (Agilent Technologies, Santa Clara, CA): the background-corrected (values < 0.01 were set to 0.01) fluorescent signal values were normalized using the Lowess method, fold-change transcript lists were generated for each microarray, and Venn diagram-based approaches were used to identify transcripts that were reproducibly informative (i.e., genes that were > 2-fold lower or higher expressed in 50% hatch compared with eye-up stage embryos on all three slides of the microarray experiment including the dye-swap).

FASTA sequences obtained from the NCBI expressed sequence tag (EST) database (Web Reference 2.6) for all accession numbers on each microarray-identified gene list, as well as for four Atlantic salmon *α-sHSP* paralogous contigs (see section 2.3.7), were aligned against NCBI's non-redundant (nr) amino acid sequence database using BLASTx (Altschul 1990; Web Reference 2.1). The best (i.e., lowest E-value) BLASTx hit with an E-value less than 1e-10 and an informative gene or protein name (i.e., disregarding 'hypothetical' or 'predicted' entries) was obtained for each feature. If a significant BLASTx hit was not retrieved for a given FASTA sequence, the sequence was aligned against NCBI's nucleotide (nt) sequence database using BLASTn (Web Reference 2.1). The best BLASTx or BLASTn hit for each microarray-identified feature and *α-sHSP* paralogue was used to query the UniProt Knowledgebase (Web Reference 2.9) to obtain putative functional

annotations (gene ontology (GO) terms) belonging to biological process, molecular function, and/or cellular component GO categories.

2.3.7 *In silico* identification of Atlantic salmon α-sHSP sequences

The cGRASP BLASTn server (Web Reference 2.1) was used to mine the cGRASP and GenBank expressed sequence tag (EST) databases for Atlantic salmon putative orthologues of four previously identified rainbow trout *α-sHSP-like* transcripts (Rise and Devlin, unpublished); this analysis identified the following Atlantic salmon contiguous sequences (contigs): cons14647, cons33468, cons69807 and cons81495. Collection and assembly of the ESTs contributing to each contig were performed using the cGRASP EST database and confirmed using the sequence assembly program CAP3 (Web References 2.2 and 2.3).

2.3.8 Sequence analysis

Hypothetical amino acid sequences and translations of putative open reading frames (ORFs) were obtained for the four microarray-identified *γM-crystallin-like* transcripts and the four *α-sHSP* paralogous contigs using the Translate Tool on the ExPASy Proteomics server (Web Reference 2.5). The correct reading frame was identified by selecting the sequence with a methionine-encoding putative start codon, and that contained no stop codon interruptions within the sequence. Translations were verified by querying the hypothetical amino acid sequences against the GenBank nr amino acid sequence database using BLASTp (Web Reference 2.1);

translations were considered correct if α-sHSP-like or γM-crystallin-like proteins were retrieved as significant hits. Additionally, confidence in the translation was increased when the length of the putative translation was in general agreement with the lengths of the top BLASTp hits. BLASTp alignment of the correct translation against NCBI's nr database also allowed identification of putative encoded domains (Web References 2.1 and 2.4). Using the default settings, MUSCLE (Edgar 2004, Web Reference 2.7) was used to create multiple sequence alignments (MSA) for: 1) four microarray-identified Atlantic salmon γM-crystallin complete or partial hypothetical amino acid sequences; 2) four Atlantic salmon and four rainbow trout α-sHSP complete hypothetical amino acid sequences. MSAs were graphically represented in Jalview (Waterhouse *et al.* 2009) to visually identify the putative γ-crystallin domain in the γM-crystallin sequences, and the putative position of the α-crystallin domain in the α-sHSP sequences, as determined by BLASTp and UniProtKb. Percent identities among suites of paralogues (i.e., the four Atlantic salmon γM-crystallins and the four Atlantic salmon α-sHSPs) were obtained by combinations of pairwise alignments using NCBI's Align/BLAST2seq feature (Web Reference 2.1). Lasergene 9 (DNASTAR, Madison, WI) was used to create phylogenetic trees (unrooted, using the neighbour joining method, bootstrapped 1000 times, random seed = 111, evolutionary distance represented by amino acid substitutions/100 residues) to determine the evolutionary relationships between microarray-identified γM-crystallin paralogues and their closest associated zebrafish

γM-crystallins or between Atlantic salmon and rainbow trout α-sHSPs.

2.3.9 QPCR primer design and quality testing

Primer3 (Web Reference 2.8) was used to design paralogue-specific primers for each Atlantic salmon *γM-crystallin* and *α-sHSP* paralogue selected for QPCR analysis. Primers were designed using the Primer3 default settings, with the following changes: primer size was set to 20 bases, and primers were designed to target areas of dissimilarity between paralogous contig sequences to ensure that only the intended paralogue was amplified. A graphical representation was also created in Jalview (Waterhouse *et al.* 2009) to indicate the location of the paralogue-specific primer sets among the *γM-crystallin* and *α-sHSP* sequences and their associated ESTs (data not shown). All candidate primer sets were quality tested following the methods in Rise *et al.* (2008). cDNA templates for QPCR were synthesized by reverse transcription from a 1 µg pool of DNase 1- treated, column-purified RNA (consisting of three high-quality 50% hatch individual samples contributing equal RNA quantities to the pool). These cDNAs were synthesized using Moloney murine leukemia virus reverse transcriptase (MMLV-RT, 200 units) (Invitrogen), 12.5 ng•µl^{-1} of random hexamers (Invitrogen), 0.5 mM of deoxyribonucleotide triphosphates (dNTPs) (Invitrogen), first strand buffer (Invitrogen), and dithiothreitol (DTT, 0.01 M) (Invitrogen) in a final volume of 20 µl (Rise *et al.* 2008). Samples were incubated in a thermal cycler (with heated lid) in the following sequence of conditions: 25 °C for 10 min, 37 °C for 50 min, 70 °C for 15 min. Finally, five-point 1:3 serial dilution

(starting with 10x dilute cDNA sample) standard curves were run for each primer set in triplicate wells using the 7500 Fast Real-Time PCR System (Applied Biosystems 7500 Fast 2.0). The QPCR reactions contained 2 µl of 10X dilute cDNA (diluted in nuclease-free water to a final volume of 200 µl), representing 10 ng of input RNA from an individual, 50 nM of the forward and reverse primers of a set, and 1X Power SYBR Green Master Mix (Applied Biosystems). Reactions were carried out using the following sequence of incubation conditions: one 10 min cycle at 95 °C, 40 cycles of 15 sec. at 95 °C and 1 min at 60 °C, with fluorescence signal measurements after each 60 °C stage. Primer sets that passed quality testing conformed to the following criteria: 1) approximately uniform spacing between the five amplification curves in the dilution series; 2) an R^2 value for the standard curve between 0.95 and 1.00, indicating a linear relationship between the serial dilutions; 3) amplification efficiency (Pfaffl 2001) between 80 and 100% (with one exception, see Table 2.1 footnotes) ; 4) low variance among technical replicates; and 5) a single peak in the melt curve, indicating the absence of primer dimers and non-specific products. Amplicons were electrophoretically separated alongside a 1 Kb Plus DNA ladder (Invitrogen) on 2.0% agarose gels to confirm that fragments of the correct size were amplified.

QPCR primer quality testing was also conducted, as previously described, for several candidate normalizer genes. In addition to the previous criteria, normalizer sets that passed quality testing had a threshold cycle (C_t) range of < 0.5 cycles for a

study, as determined by testing all individuals across all time points. Although six microarray-identified candidate normalizers (i.e., background corrected, Lowess normalized expression ratio between 0.8 and 1.2 fold on all slides of the microarray study, including the dye-swap) were tested, none passed quality testing (i.e., none had a C_t range of < 0.5 cycles) (data not shown). Subsequently, primer sets for four additional candidate normalizers commonly used in the Rise Lab were also quality tested (data not shown). Paralogue-specific *γM-crystallin* and *α-sHSP* primer sets, and the Atlantic salmon *polyadenylate (poly-A) binding protein 1* (*PABP1*) normalizer primer set that passed quality testing and was used in QPCR analyses, are summarized in Table 2.1.

2.3.10 cDNA synthesis and QPCR assays of selected genes of interest

cDNA for each individual sample was synthesized (Rise *et al.* 2008) from 1 µg of high-quality, DNase 1-treated and column-purified RNA, using MMLV-RT (Invitrogen) as per the manufacturer's instructions, as described above (section 2.3.8.). The cDNA samples were diluted 10X in nuclease free water and kept at-20 °C until use. In addition, a cDNA pool of all individuals in the study, generated using 1 µg of DNase 1-digested and column-purified RNA per individual (RNA pooled prior to cDNA synthesis) was used to run "no-reverse transcription" control assays (i.e., omitting the reverse transcriptase in cDNA synthesis). All QPCR was run with technical triplicate reactions for each primer set on the 7500 Fast Real-Time PCR

Table 2.1 Paralogue-specific primer sets used for QPCR analysis of four paralogous *γM-crystallin* transcripts and four paralogous *α-sHSP* transcripts in Atlantic salmon, along with those for the normalizer gene (*PABP1*).

Gene of Interest	Primer Name	Primer Length (bases)	Sequence (5' – 3')	Amplicon Size (bp)	Amplification Efficiency (%)[1]
***S.salar* cryGM2-like 1 (cons47274)**	1cryGM2-1_L	20	CCCATGATGTCGTCACAGTC	153	81.8
	1cryGM2-1_R	20	CTCTGACTACCAGCGCATGA		
***S.salar* cryGM2-like 2 (cons68532)**	4cryGM2-2_L	20	CATGAACCTCATGCCACTCA	160	92.0
	4cryGM2-2_R	20	CCAGGAGCGTTACCGTATGT		
***S.salar* cryGM4-like (cons176022)**	3cryGM4_L	20	TGTACATCTGCCTGCCTCTG	215	85.6
	3cryGM4_R	20	CATCAGGTCTTGCCGTAACA		
***S.salar* cryGM3-like (cons14521)**	4cryGM3_L	20	AGTCGCTCATGGTCATACCC	168	68.4[2]
	4cryGM3_R	20	TCGTTCCTATGAGACCAGCA		
***S.salar* α-sHSP 1**	2sHSP1_L	20	AAACAGGCCAAAGCAGAAGA	115	95.1
	2sHSP1_R	20	GAGCAGGCAGGTCAATTTTC		
***S.salar* α-sHSP 2**	2sHSP2_L	20	AAACAGGCCAAAGCAGAAGA	141	101.9
	2sHSP2_R	20	ACAGGTCAAGGCTGTCAGGT		
***S.salar* α-sHSP 3**	8sHSP3_L	20	CCACCTGCTTGACAGACAGA	164	87.9
	8sHSP3_R	20	TATGCAACGACTCCACCAAA		
***S.salar* α-sHSP 4**	1sHSP4_L	20	ATATGCAACGACTGCACCAA	148	107.9
	1sHSP4_R	20	AGCTCCTCTGGGGAGAAGTC		
Poly-A binding protein 1 (normalizer[3])	PABP 1L	20	TGACCGTCTCGGGTTTTTAG	108	101.8
	PABP 1R	20	CCAAGGTGGATGAAGCTGTT		

[1]% Efficiency values (from standard curves) were used to calculate relative quantity (RQ).

[2]Three paralogue-specific primer sets were designed and quality tested for *S. salar cryGM3-like*, and this set had the best quality characteristics. Actual primer efficiency values were utilized in the RQ calculations (see Methods).

[3]Although ten candidate normalizers were tested, *PABP1* had the most stable transcript expression (all C_t values within a range of < 0.5 cycles) across all time points and individuals in the study.

System (Applied Biosystems 7500 Fast 2.0). No amplification was observed in the "no-reverse transcription" controls, confirming the absence of genomic DNA in the RNA templates. QPCR analysis was used to determine transcript expression and individual biological variability using six biological replicates for each time point (i.e., two individuals from each of three tanks per time point), and three technical replicates for each individual and gene of interest. Fluorescence and baseline thresholds were automatically set by the 7500 Fast Real-Time PCR system. The expression of the eight GOI (four microarray-identified *γM-crystallin* transcripts and four *α-sHSP* paralogous transcripts) was normalized to the expression of *PABP1*. A control amplification reaction was run on each plate of the study, in order to link plates and evaluate technical variability. The linker sample (a cDNA template synthesized by reverse transcription from a 1 μg pool of DNase 1-treated, column-purified RNA, consisting of equal contributions from three high-quality individual samples from the eye-up time point) was included for each gene of interest as well as the normalizer, with an inter-plate C_t range of < 0.5 cycles. C_t values were obtained and used to calculate the relative quantity (RQ) of each transcript using the $2^{-\Delta\Delta Ct}$ quantification method (Livak and Schmittgen 2001; Pfaffl 2001) and the Applied Biosystems (Foster City, CA) 7500 Software Relative Quantification Study Application (v2.0) (Livak and Schmittgen 2001). The specific amplification efficiency for each primer set (see Table 2.1), which was determined during quality testing, was incorporated in the calculation of each RQ value.

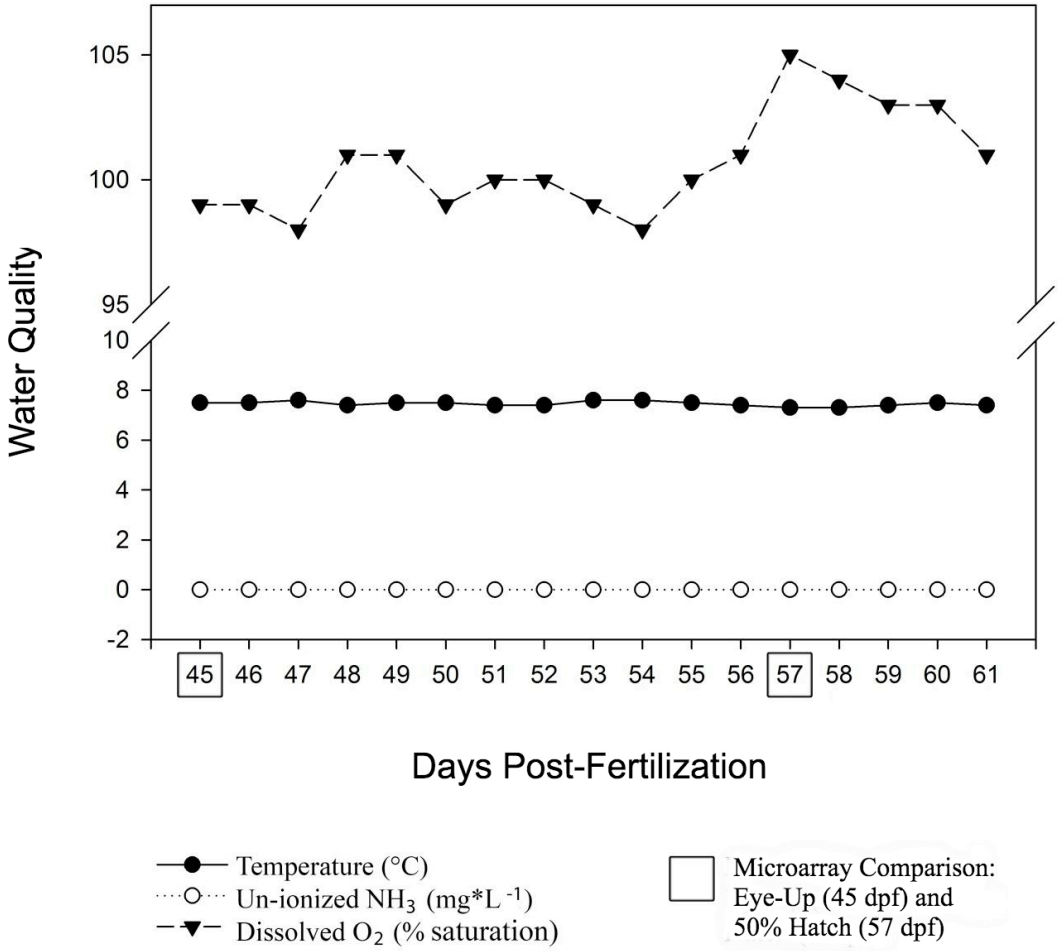
RQ values are expressed as mean values ± SEM. The individual with the lowest normalized expression for a particular target (regardless of the sampling time point of that individual) was used as the calibrator (i.e., RQ set to 1.0). Overall fold change values were calculated relative to the initial sampling point (eye-up, at 45 dpf); for example, fold change for a GOI at 57 dpf was calculated as (mean RQ for 57 dpf samples)/(mean RQ for 45 dpf samples), representing overall fold change at 57 dpf compared to 45 dpf. Statistical differences (SigmaPlot 12 for Windows) were determined using a one-way analysis of variance (ANOVA), followed by Tukey post-hoc tests, and considered significant if $p < 0.05$.

2.4 RESULTS

2.4.1 Water quality monitoring

No significant differences in water quality were observed over the course of the study (i.e., one-way repeated measures ANOVA, $p < 0.05$). Water temperature in the tanks was very stable over the course of the study, averaging (± SEM) 7.32 ± 0.04 °C. NH_3 levels were consistently low, with an average of 0.00005 ± 0.000002 mg•L^{-1}, while DO averaged 98.9 ± 0.24% (Figure 2.2).

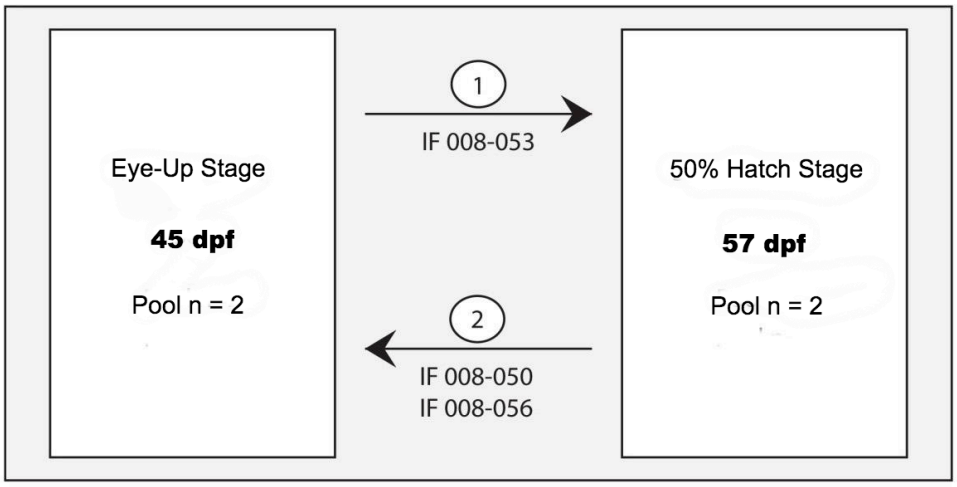
Figure 2.2 Water quality during Atlantic salmon development from eye-up stage (45 dpf) to post-hatch (61 dpf). Water was tested every 24 h for ambient temperature (°C), un-ionized ammonia (NH_3) concentration ($mg \bullet L^{-1}$), and dissolved oxygen saturation (% O_2).


Water Quality
105
100
95
10
8
6
4
2
0
-2
45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61
Days Post-Fertilization
Temperature (°C)
Un-ionized NH3 (mg*L -1)
Dissolved O2 (% saturation)
Microarray Comparison:
Eye-Up (45 dpf) and
50% Hatch (57 dpf)

2.4.2 Microarray hybridization using the 32K cGRASP salmonid array platform, and identification of Atlantic salmon *γM-crystallin* transcripts

The 32K cGRASP salmonid cDNA microarray was used in a direct comparison of global gene expression between Atlantic salmon at eye-up (45 dpf) and 50% hatch (57 dpf) developmental stages (Figure 2.3). Transcripts that were greater than 2-fold lower expressed in 50% hatch compared to eye-up stage embryos were represented by 11 microarray features (Table 2.2). Transcripts that were greater than 2-fold more highly expressed in 50% hatch compared to eye-up stage embryos were represented by 73 microarray features (Table 2.3). In order to identify developmentally regulated putative paralogues for subsequent QPCR studies (i.e., to look for evidence of gene duplication and divergence), gene names that were represented multiple times in a given gene list were noted. Although no evidence of paralogues (i.e., multiple same-named features) was found in the list of 11 microarray features that were lower expressed in 50% hatch embryos (Table 2.2), the gene list of 73 microarray features that were more highly expressed in hatch stage embryos had several gene or protein names that were represented multiple times (Table 2.3). For example, Table 2.3 contains eight *γM-crystallin* –like features, twelve *parvalbumin-like* features, and four features identified as *glyceraldehyde-3-phosphate dehydrogenase*. The *γ-crystallins* are known to be an expanded gene family, and have been previously studied as a model of gene family evolution (e.g.,

Figure 2.3 Overview of the experimental design using the cGRASP 32K salmonid microarray to compare global gene expression between embryos at eye-up (45 dpf) and 50% hatch (57 dpf) developmental stages. Targets were fluorescently labeled with either Cy3 or Cy5 (see Methods for details). The batch and serial number of the array, as well as the numbers of genes that passed threshold and were reproducibly greater than 2-fold lower or higher expressed in the 50% hatch embryonic RNA pool (n = 2: one individual from two of the triplicate tanks) compared with the eye-up embryonic RNA pool (n = 2: one individual from two of the triplicate tanks) on all three technical replicate arrays, including one dye-swap (see Tables 2.5 and 2.6), are indicated. Numbers above the arrows indicate the number of technical replicate arrays. See section 2.3.5 for details about the microarray design, including reasoning on the selection of individuals for the RNA pools.



The expression of 11 genes was > 2-fold lower expressed in embryos at 50% hatch (57 dpf) compared to eye-up (45 dpf) stage embryos (Table 2.2)

The expression of 73 genes was > 2-fold higher expressed in embryos at 50% hatch (57 dpf) compared to eye-up (45 dpf) stage embryos (Table 2.3)

Table 2.2 Eleven reproducibly informative[1] microarray features with greater than 2-fold lower expression in Atlantic salmon 50% hatch (57 dpf) stage embryos compared to eye-up (45 dpf) stage embryos.

EST Accession Number	Gene Name of Best[2] BLAST[3] Hit	Length (%)[4]	E-value	Functional Annotation[5] of Best BLAST Hit	GO Identifier[6]	Mean Fold Change	SEM[7]
EG869972	Transcriptome shotgun assembly (TSA): 28706 mRNA sequence (EZ792040 ; *Oncorhynchus mykiss*)[3]	664 (87)	0	N/F	N/F	5.72	0.90
EG888926	Unknown	N/A	N/A	N/F	N/F	4.03	0.97
CA042975	Type II keratin E3 (CAC87008; *Oncorhynchus mykiss*)	231 (98)	4e-119	MF: structural molecule activity; CC: keratin filament, intermediate filament.[8]	GO:0005198 GO:0045095 GO:0005882	3.30	0.96
DY696793	Unknown	N/A	N/A	N/F	N/F	3.22	0.92
DY703407	Upf0553 protein c9orf64-like protein (NP_001187367; *Ictalurus punctatus*)	76 (66)	1e-21	N/F	N/F	3.19	0.22
CA062798	Unknown	N/A	N/A	N/F	N/F	3.18	1.02
DW563998	TSA: 30549 mRNA sequence (EZ793883; *Oncorhynchus mykiss*)[3]	539 (93)	0	N/F	N/F	2.94	0.41
CK990883	Embryonic beta-type globin2 (BAA34951; *Oncorhynchus mykiss*)	131 (88)	1e-59	BP: oxygen transport, transport; MF: iron ion binding, metal ion binding, heme binding, oxygen transporter activity, oxygen binding; CC: hemoglobin complex.[8]	GO:0015671 GO:0006810 GO:0005506 GO:0046872 GO:0020037 GO:0005344 GO:0019825 GO:0005833	2.83	0.31

EG912741	Cathepsin F (ACI33686; *Salmo salar*)	475 (99)	0	BP: proteolysis; MF: cysteine-type peptidase activity, peptidase activity, cysteine-type endopeptidase inhibitor activity, hydrolase activity, cysteine-type endopeptidase activity.	GO:0006508 GO:0008234 GO:0008233 GO:0004869 GO:0016787 GO:0004197	2.71	0.39
DY712654	TSA: 72742 mRNA sequence (EZ836076; *Oncorhynchus nerka*)[3]	462 (85)	3e-121	N/F	N/F	2.65	0.27
CX246745	Unknown	N/A	N/A	N/F	N/F	2.20	0.02

[1]"Reproducibly informative" is defined as ≥ 2-fold higher expressed in eye-up (45 dpf) compared to 50% hatch (57 dpf) stage embryos on all three technical replicate microarrays including one dye-swap.

[2]The best BLASTx or BLASTn hit, defined as having the lowest E-value (≤ 1e-10) and an associated protein or gene name (i.e., not "hypothetical" or "predicted"), is shown. Species: *Ictalurus punctatus* (channel catfish); *Oncorhynchus mykiss* (rainbow trout); *Oncorhynchus nerka* (sockeye salmon); and *Salmo salar* (Atlantic salmon). BLAST reports were collected on May 27th, 2011, and reflect the state of the GenBank non-redundant (nr) amino acid database and nucleotide (nt) database on this date. N/A: not applicable.

[3]Indicates that a BLASTn hit was used where no significant BLASTx hit was retrieved.

[4]Length of aligned region, in amino acid residues, and the percent identity, indicating the number of exactly matched residues between the query sequence (i.e., the microarray-identified feature) and its best BLAST hit.

[5]Functional annotation associated with *Salmo salar*. N/F: no functional annotation found for best BLAST hit or any putative orthologues. Gene ontology (GO) categories: biological process (BP), molecular function (MF) and cellular component (CC).

[6]GO identifiers are listed in the same relative order as the associated functional annotations (GO terms).

[7]Since pooled templates were used in this experiment, SEM values carry information on technical variability, rather than biological variability. However, the QPCR experiments used individual samples rather than pools.

[8]Functional annotations associated with *Oncorhynchus mykiss*.

Table 2.3 Seventy-three reproducibly informative[1] microarray features with greater than 2-fold higher expression in Atlantic salmon 50% hatch (57 dpf) stage embryos compared to eye-up (45 dpf) stage embryos. *γM-crystallin-like* transcripts are indicated in grey.

EST Accession Number	**Gene Name of Best[2] BLAST[3] Hit**	**Length (%)[4]**	**E-value**	**Functional Annotation[5] of Best BLAST Hit**	**GO Identifier[6]**	**Mean Fold Change**	**SEM[7]**
CB496718	Parvalbumin beta 27 (AAN10127; *Salvelinus alpinus*)	107 (93)	7e -37	MF: calcium ion binding.[8]	GO:0005509	531.99	515.61
CB510226	Parvalbumin beta 27 (AAN10127; *Salvelinus alpinus*)	107 (95)	4e -43	MF: calcium ion binding.[8]	GO:0005509	19.50	3.53
CB510202	Parvalbumin-2 (ACH70862; *Salmo salar*)	109 (100)	1e -48	MF: calcium ion binding.	GO:0005509	16.98	9.35
CB509945	**Crystallin gamma M2b (ACL98110; *Epinephelus coioides*)**	**104 (83)**	**9e -49**	**N/F**	**N/F**	**14.56**	**7.51**
	Crystallin gamma M2c (NP_001007784.1; *Danio rerio*)	**106 (78)**	**1e -58**	**N/F**	**N/F**		
CA038638	Gamma M2c crystallin (ABA61352; *Dissostichus mawsoni*)	173 (71)	3e -74	N/F	N/F	11.39	5.68
	Crystallin gamma M2b (NP_001018619.1; *Danio rerio*)	172 (68)	6e -89	N/F	N/F		
CB510792	**Gamma M2b crystallin (ABA61351; *Dissostichus mawsoni*)**	**179 (75)**	**1e -82**	**N/F**	**N/F**	**11.36**	**6.04**
	Crystallin gamma M4 (NP_001007792.1; *Danio rerio*)	**178 (70)**	**8e -85**	**N/F**	**N/F**		

CB517835	Tropomyosin-1 alpha chain putative mRNA, clone ssal-rgf-518-106 complete cds (BT045316; *Salmo salar*)[3]	602 (99)	0	N/F	N/F	10.14	0.51
CB496509	Parvalbumin beta 27 (AAN10127; *Salvelinus alpinus*)	107 (93)	7e -37	MF: calcium ion binding.[8]	GO:0005509	9.49	1.54
CB496985	Troponin C fast (ACH70760; *Salmo salar*)	61 (96)	2e -26	MF: calcium ion binding.	GO:0005509	8.37	2.26
CB496355	Parvalbumin (CAX32967; *Salvelinus fontinalis*)	87 (98)	4e -41	MF: calcium ion binding.[8]	GO:0005509	8.11	2.93
CB510426	Unknown	N/A	N/A	N/F	N/F	7.87	2.64
BU965756	Glyceraldehyde-3-phosphate dehydrogenase (ACI69846; *Salmo salar*)	116 (100)	2e -60	BP: glucose metabolic process, glycolysis, oxidation reduction; MF: NAD or NADH binding, glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) activity, glyceraldehyde-3-phosphate dehydrogenase activity, oxidoreductase activity; CC: cytoplasm.	GO:0006006 GO:0006096 GO:0055114 GO:0051287 GO:0004365 GO:0008943 GO:0016491 GO:0005737	7.31	1.57
CB496584	Alpha actin (AF267496; *Salmo trutta*)	112 (99)	7e -59	MF: nucleotide binding, ATP binding, protein binding; CC: cytoplasm, cytoskeleton.[8]	GO:0000166 GO:0005524 GO:0005515 GO:0005737 GO:0005856	7.17	1.01
CB493676	Gamma M2c crystallin (ABA61352; *Dissostichus mawsoni*)	168 (76)	3e -78	N/F	N/F	6.87	1.71

	Crystallin, gamma M2f (NP_001103576.1; *Danio rerio*)	168 (70)	6e -90	N/F	N/F		
CB511383	Slow myosin light chain 2 (ABU97483; *Oncorhynchus mykiss*)	156 (98)	6e -85	MF: calcium ion binding.[9]	GO:0005509	6.70	1.96
CB494699	Crystallin gamma M2b (ACL98110; *Epinephelus coioides*)	129 (83)	2e -64	N/F	N/F	6.17	0.48
	Crystallin gamma M2f (NP_001103576.1; *Danio rerio*)	130 (81)	9e -89	N/F	N/F		
CB491826	Glyceraldehyde-3-phosphate dehydrogenase (ACI69846; *Salmo salar*)	162 (99)	1e -85	BP: glucose metabolic process, glycolysis, oxidation reduction; MF: NAD or NADH binding, glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) activity, glyceraldehyde-3-phosphate dehydrogenase activity, oxidoreductase activity; CC: cytoplasm.	GO:0006006 GO:0006096 GO:0055114 GO:0051287 GO:0004365 GO:0008943 GO:0016491 GO:0005737	5.98	1.72
CB496872	Parvalbumin (CAX32967; *Salvelinus fontinalis*)	87 (98)	4e -41	MF: calcium ion binding.[8]	GO:0005509	5.41	1.08
CB511371	**Gamma M5a crystallin (ABA61353; *Dissostichus mawsoni*)**	**173 (77)**	**1e -79**	**N/F**	**N/F**	**5.38**	**0.45**
	Crystallin gamma M3 (NP_001007787.1; *Danio rerio*)	**173 (69)**	**2e -86**	**N/F**	**N/F**		

CA051900	Cathepsin L1 (NP_001140018; *Salmo salar*)	144 (100)	3e -82	BP: proteolysis; MF: peptidase activity, hydrolase activity, cysteine-type peptidase activity, cysteine-type endopeptidase activity.	GO:0006508 GO:0008233 GO:0016787 GO:0008234 GO:0004197	5.29	1.17
CB514460	Glyceraldehyde-3-phosphate dehydrogenase (ACI69846; *Salmo salar*)	200 (100)	2e -110	BP: glucose metabolic process, glycolysis, oxidation reduction; MF: NAD or NADH binding, glycerald-ehyde-3-phosphate dehydrogenase (phosphorylating) activity, glycer-aldehyde-3-phosphate dehydrogenase activity, oxidoreductase activity; CC: cytoplasm.	GO:0006006 GO:0006096 GO:0055114 GO:0051287 GO:0004365 GO:0008943 GO:0016491 GO:0005737	5.24	0.70
CB498646	Myosin light chain 1-1 (ACH71036; *Salmo salar*)	65 (98)	3e -28	MF: calcium ion binding.	GO:0005509	5.22	0.69
CB510589	**Crystallin gamma M2b (ACL98110; *Epinephelus coioides*)**	**183 (81)**	**1e -92**	**N/F**	**N/F**	**5.19**	**0.22**
	Crystallin, gamma M2f (NP_001103576.1; *Danio rerio*)	**175 (78)**	**1e -102**	**N/F**	**N/F**		
CB509748	Parvalbumin beta 2 (CAA66404; *Salmo salar*)	87 (100)	2e -41	MF: calcium ion binding.	GO:0005509	4.98	0.48

CB493716	IS10 transposase (NP_058298; *Salmonella typhi*)	205 (100)	4e -114	BP: transposition, DNA-mediated; MF: DNA binding, nucleic acid binding, transposase activity.[10]	GO:0006313 GO:0003677 GO:0003676 GO:0004803	4.92	0.98
CB494797	Crystallin gamma M2b (ACL98110; *Epinephelus coioides*)	70 (85)	3e -43	N/F	N/F	4.75	0.65
	Crystallin, gamma M2a (NP_001018131.1; *Danio rerio*)	70 (86)	4e - 40	N/F	N/F		
EG774444	Actin (AAU94670; *Amoebidium parasiticum*)	141 (93)	3e -72	MF: nucleotide binding, ATP binding, protein binding; CC: cytoplasm, cytoskeleton.[10]	GO:0000166 GO:0005524 GO:0005515 GO:0005737 GO:0005856	4.65	2.00
CB505581	Skeletal alpha-actin 2 protein (ADD81348; *Cobitis choii*)	42 (100)	8e -17	MF: protein binding, ATP binding, nucleotide binding; CC: cytoplasm, cytoskeleton.[8]	GO:0005515 GO:0005524 GO:0000166 GO:0005737 GO:0005856	4.60	0.23
EG784759	Troponin C fast (ACH70760; *Salmo salar*)	160 (100)	1e -76	MF: calcium ion binding.	GO:0005509	4.59	1.15
CB496977	Fast myotomal muscle actin 2 (ACH70714; *Salmo salar*)	192 (100)	2e -105	MF: protein binding, ATP binding, nucleotide binding; CC: cytoplasm, cytoskeleton.	GO:0005515 GO:0005524 GO:0000166 GO:0005737 GO:0005856	4.58	0.60
CA064436	Tropomyosin-1 alpha chain putative mRNA, clone ssal-rgf-536-260 complete cds (BT045917; *Salmo salar*)[3]	755 (99)	0	N/F	N/F	4.52	0.56

CB494389	Creatine kinase (ACH70839; *Salmo salar*)	102 (99)	3e -40	MF: transferase activity, transferring phosphorus-containing groups, ATP binding, kinase activity, transferase activity, nucleotide binding, catalytic activity.	GO:0016772 GO:0003824 GO:0016301 GO:0005524 GO:0016740 GO:0000166	4.28	0.83
CB494403	Creatine kinase-2 (ACH70914; *Salmo salar*)	176 (100)	1e -100	MF: transferase activity, transferring phosphorus-containing groups, ATP binding, kinase activity, transferase activity, nucleotide binding, catalytic activity.	GO:0016772 GO:0003824 GO:0016301 GO:0005524 GO:0016740 GO:0000166	4.28	0.79
CB494506	Parvalbumin (CAX32967; *Salvelinus fontinalis*)	87 (100)	8e -42	MF: calcium ion binding.[8]	GO:0005509	4.24	0.63
CA040114	Tropomyosin-1 alpha chain putative mRNA, clone ssal-rgf-536-260 complete cds (BT045917; *Salmo salar*)[3]	707 (99)	0	N/F	N/F	4.22	1.30
CB497148	Unknown	N/A	N/A	N/F	N/F	4.15	0.95
CB510271	Parvalbumin beta 2 (CAA66404; *Salmo salar*)	108 (100)	1e -43	MF: calcium ion binding.	GO:0005509	4.14	0.87
CA053670	Protein kinase C substrate 80K-H (AAH46883; *Danio rerio*)	42 (78)	6e -13	BP: pronephros development; MF: protein binding, kinase activity.[11]	GO:0048793 GO:0005515 GO:0016301	4.10	0.93

EG898703	Beta actin isoform 1 (BAG12566; *Solea senegalensis*)	244 (95)	2e -130	MF: ATP binding, nucleotide binding, protein binding; CC: cytoskeleton, cytoplasm.[8]	GO:0005524 GO:0000166 GO:0005515 GO:0005856 GO:0005737	4.07	1.24
CB494208	Creatine kinase (ACH70839; *Salmo salar*)	107 (100)	6e -144	MF: transferase activity, transferring phosphorus-containing groups, catalytic activity, kinase activity, ATP binding, nucleotide binding.	GO:0016772 GO:0003824 GO:0016301 GO:0005524 GO:0016740 GO:0000166	4.04	0.37
CB510421	Deoxyribo-nuclease gamma precursor (ACI69646; *Salmo salar*)	225 (98)	1e -91	BP: DNA catabolic process; MF: deoxyribonuclease activity.	GO:0006308 GO:0004536	4.01	0.85
CB511208	Fast myotomal muscle actin (AAG25672; *Salmo salar*)	138 (100)	8e -75	MF: ATP binding, nucleotide binding, protein binding; CC: cytoskeleton, cytoplasm.	GO:0005524 GO:0000166 GO:0005515 GO:0005856 GO:0005737	3.93	0.25
CB497026	Cathepsin L1 precursor (ACO15730; *Caligus clemens*i)	131 (100)	1e -73	BP: proteolysis; MF: cysteine-type peptidase activity, cysteine-type endopeptidase activity, peptidase activity, hydrolase activity.[10]	GO:0006508 GO:0008234 GO:0004197 GO:0008233 GO:0016787	3.88	1.06
CK990375	Unknown	N/A	N/A	N/F	N/F	3.81	0.38
CA056748	Unknown	N/A	N/A	N/F	N/F	3.81	0.66
CK990562	Cathepsin L1 precursor (CAN10121; *Salmo salar*)	92 (95)	1e -46	BP: proteolysis; MF: peptidase activity, hydrolase activity, cysteine-type peptidase activity, cysteine-type endopeptidase activity.	GO:0006508 GO:0008233 GO:0016787 GO:0008234 GO:0004197	3.76	0.70
CK990679	Unknown	N/A	N/A	N/F	N/F	3.73	0.80

CB509752	Parvalbumin (CAX32967; *Salvelinus fontinalis*)	56 (100)	2e -23	MF: calcium ion binding.[8]	GO:0005509	3.71	0.42
CB496670	Gelsolin precursor putative mRNA, clone ssal-rgf-511-021 partial cds (BT072088 *Salmo salar*)[3]	614 (88)	0	MF: actin binding.	GO:0003779	3.69	0.73
CB492406	Parvalbumin (CAX32967; *Salvelinus fontinalis*)	87 (71)	4e -21	MF: calcium ion binding.[8]	GO:0005509	3.69	0.71
CB493415	Myosin regulatory light chain 2 (CAD89610; *Salmo salar*)	141 (98)	6e -74	MF: calcium ion binding.	GO:0005509	3.65	0.52
CB509702	Parvalbumin (CAX32967; *Salvelinus fontinalis*)	31 (100)	0	MF: calcium ion binding.[8]	GO:0005509	3.55	0.36
CB498361	Glyceraldehyde-3-phosphate dehydrogenase (ACI66269; *Salmo salar*)	146 (99)	6e -78	BP: oxidation reduction, glucose metabolic process, glycolysis; MF: NAD or NADH binding, oxidoreductase activity, glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) activity, glyceraldehyde-3-phosphate dehydrogenase activity; CC: cytoplasm.	GO:0055114 GO:0006006 GO:0006096 GO:0051287 GO:0016491 GO:0004365 GO:0008943 GO:0005737	3.52	0.23
CB496867	60S ribosomal protein L10 (ACN10030; *Salmo salar*)	150 (99)	3e -82	BP: translation; MF: structural constituent of ribosome; CC: intracellular, ribosome.	GO:0006412 GO:0003735 GO:0005622 GO:0005840	3.45	1.14

CB510978	Fast myotomal muscle tropomyosin (AAB36559; *Salmo salar*)	172 (100)	1e -87	N/F	N/F	3.40	0.47
CK991305	Non-metastatic cells 1 protein (ACH70800; *Salmo salar*)	142 (92)	2e -72	BP: CTP biosynthetic process, GTP biosynthetic process, UTP biosynthetic process; MF: ATP binding, nucleotide binding, nucleoside diphosphate kinase activity, transferase activity, kinase activity; CC: cytoplasm.	GO:0006241 GO:0006183 GO:0006228 GO:0005524 GO:0000166 GO:0004550 GO:0016740 GO:0016301 GO:0005737	3.37	0.44
CB491722	Non-metastatic cells 1 protein (NP_001133149; *Salmo salar*)	135 (96)	7e -71	BP: CTP biosynthetic process, GTP biosynthetic process, UTP biosynthetic process; MF: ATP binding, nucleotide binding, nucleoside diphosphate kinase activity, transferase activity, kinase activity; CC: cytoplasm.	GO:0006241 GO:0006183 GO:0006228 GO:0005524 GO:0000166 GO:0004550 GO:0016740 GO:0016301 GO:0005737	3.26	0.71
CK991104	Unknown	N/A	N/A	N/F	N/F	3.23	0.17
CA036990	Cathepsin L-like protein (ACO82386; *Lutjanus argentimaculatus*)	87 (91)	3e -43	BP: proteolysis; MF: peptidase activity, hydrolase activity, cysteine-type peptidase activity, cysteine-type endopeptidase activity.[8]	GO:0006508 GO:0008233 GO:0016787 GO:0008234 GO:0004197	3.20	0.87

CB510287	Myosin, light polypeptide 3-3 (ACH70957; *Salmo salar*)	105 (100)	3e -51	MF: calcium ion binding.	GO:0005509	3.16	0.41
EG883987	Aatf protein (AAI16504; *Danio rerio*)	186 (65)	3e -34	CC: nucleus.[11]	GO:0005634	3.13	0.92
CA057166	Enolase 3-2 (NP_001133193; *Salmo salar*)[3]	167 (99)	8e -92	BP: glycolysis; MF: phosphopyruvate hydratase activity, magnesium ion binding, lyase activity; CC: cell surface, phosphopyruvate hydratase complex, cytoplasm.	GO:0006096 GO:0004634 GO:0000287 GO:0016829 GO:0009986 GO:0000015 GO:0005737	3.10	0.29
CA057774	Actin 1 (BAJ13478; *Guillardia theta*)	23 (100)	0	N/F	N/F	2.97	0.27
CB514277	Skeletal alpha-actin 2 protein (ADD81348; *Cobitis choii*)	101 (100)	6e -52	MF: protein binding, ATP binding, nucleotide binding; CC: cytoplasm, cytoskeleton.[8]	GO:0005515 GO:0005524 GO:0000166 GO:0005737 GO:0005856	2.96	0.58
CB497075	RING finger protein 31 putative mRNA, clone ssal-rgf-508-247 pseudogene cds (BT072038 *Salmo salar*)[3]	340 (85)	6e -84	N/F	N/F	2.91	0.40
CB498218	Creatine kinase-3 (ACH70916; *Salmo salar*)	85 (100)	5e -30	MF: transferase activity, transferring phosphorus-containing groups, ATP binding, kinase activity, transferase activity, nucleotide binding, catalytic activity.	GO:0016772 GO:0003824 GO:0016301 GO:0005524 GO:0016740 GO:0000166	2.89	0.12

CB509968	Fast myotomal muscle troponin-T-2 (ACH71020; *Salmo salar*)	85 (100)	1e -23	N/F	N/F	2.88	0.38
CB497701	Unknown	N/A	N/A	N/F	N/F	2.77	0.35
CB493574	Glyceraldehyde-3-phosphate dehydrogenase-2 (ACH70939; *Salmo salar*)	124 (100)	5e -60	BP: glucose metabolic process, glycolysis, oxidation-reduction process; MF: NAD binding, glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity, oxidoreductase activity, nucleotide binding, oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor, NADP binding; CC: cytoplasm.	GO:0006006 GO:0006096 GO:0055114 GO:0051287 GO:0004365 GO:0016491 GO:0000166 GO:0016620 GO:0050661 GO:0005737	2.74	0.40
CA058389	Creatine kinase B-type (NP_001133250.1; *Salmo salar*)	154 (100)	4e -76	MF: transferase activity, transferring phosphorus-containing groups, catalytic activity, kinase activity, ATP binding, nucleotide binding.	GO:0016772 GO:0003824 GO:0016301 GO:0005524 GO:0016740 GO:0000166	2.64	0.08
EG841611	Aspartyl beta-hydroxylase-like (ACH70812; *Salmo salar*)	42 (97)	3e -16	MF: calcium ion binding.	GO:0005509	2.61	0.22

CK991019	Creatine kinase B-type (NP_001133250; *Salmo salar*)	128 (91)	6e -53	MF: transferase activity, transferring phosphorus-containing groups, catalytic activity, kinase activity, ATP binding, transferase activity, nucleotide binding.	GO:0016772 GO:0003824 GO:0016301 GO:0005524 GO:0016740 GO:0000166	2.61	0.24
CB497161	CD9 antigen putative mRNA, clone ssal-rgf-508-015 complete cds (BT058855; *Salmo salar*)	538 (84)	8e -137	CC: membrane, integral to membrane.	GO:0016020 GO:0016021	2.50	0.02

[1]"Reproducibly informative" is defined as ≥ 2-fold higher expressed in 50% hatch (57 dpf) compared to eye-up (45 dpf) stage embryos on all three technical replicate microarrays including one dye-swap.

[2]The best BLASTx or BLASTn hit, defined as having the lowest E-value (≤ 1e-10) and an associated protein or gene name (i.e., not "hypothetical" or "predicted"), is shown. Species: *Amoebidium parasiticum* (protozoan, no common name found); *Caligus clemensi* (sea louse); *Cobitis choii* (Choi's spiny loach); *Danio rerio* (zebrafish); *Dissostichus mawsoni* (Antarctic toothfish); *Epinephelus coioides* (orange-spotted grouper); *Guillardia theta* (cryptomonad alga, no common name found); *Lutjanus argentimaculatus* (mangrove red snapper); *Oncorhynchus mykiss* (rainbow trout); *Salmonella typhi* (parasitic typhoid bacillus, no common name found); *Salmo salar* (Atlantic salmon); *Salmo trutta* (brown trout); *Salvelinus alpinus* (Arctic char); *Salvelinus fontinalis* (brook trout); and *Solea senegalensis* (Senegalese sole). BLAST reports were collected on May 27th, 2011, and reflect the state of the GenBank non-redundant (nr) amino acid database and nucleotide (nt) database on this date. N/A: not applicable. Atlantic salmon associated *γM-crystallin-like* transcripts are indicated in grey, with bold font used to identify features that were later used in contig assembly and QPCR; the best *Danio rerio* (zebrafish) BLASTx hit with the lowest E-value (≤ 1e-10) and an associated protein or gene name (i.e., not "hypothetical" or "predicted") is also shown for each *γM-crystallin-like* transcript.

[3]Indicates that a BLASTn hit was used where no significant BLASTx hit was retrieved.

[4]Length of aligned region, in amino acid residues, and the percent identity, indicating the number of exactly matched residues between the query sequence (i.e., the microarray-identified feature) and its best BLAST hit.

[5]Functional annotation associated with *Salmo salar*. N/F: no functional annotation found for best BLAST hit or any putative orthologues. Gene ontology (GO) categories: biological process (BP), molecular function (MF) and cellular component (CC).

[6]GO identifiers are listed in the same relative order as the functional annotations (GO terms) with which they were associated. Microarray-identified genes of interest (GOI) that were further analysed by QPCR are indicated in bold.

[7]Since pooled templates were used in this experiment, SEM values carry information on technical variability, rather than biological variability. However, the QPCR experiments used individual samples rather than pools.

[8]Functional annotations associated with teleost fish, *Oncorhynchus mykiss*[9], *Danio rerio*[10], or other species[11].

Weadick and Chang 2009.; Greiling *et al.* 2009; Kappé *et al.* 2010; Slingsby *et al.* 2013). Therefore, the microarray-identified *γM-crystallin-like* features were selected for paralogue discovery and paralogue-specific QPCR, and the best *Danio rerio* BLAST hits for these features suggested that multiple paralogues (i.e., M2a, M2c, M3) were represented in Table 2.3.

2.4.3 Characterization of four paralogous Atlantic salmon *γM-crystallin* contiguous sequences

Eight Atlantic salmon *γM-crystallin-like* features were identified by microarray (Table 2.3). Multiple sequence alignment and phylogenetic tree construction using these eight salmon sequences and their zebrafish homologues showed that four of the transcripts were relatively more closely related (CB509945, CB510589, CB510792, and CB511371), whereas the other four transcripts (CA038638, CB493676, CB494699, and CB484797) were relatively more distantly related (data not shown - the relationships among the four selected transcripts is shown in Figure 2.4A). The cGRASP EST database and BLAST server (Web Reference 2.3) were used to identify contigs and contributing ESTs representing the four closest related features (Table 2.4). The first contig, named *S. salar cryGM2-like 1* (CB509945; cons47274) was 830 bases in length and comprised of eight Atlantic salmon clone sequences. *S. salar cryGM2-like 2* (CB510589; cons68532), a contig 629 bases in length, was assembled using five Atlantic salmon clones. Ten clones were assembled in the cGRASP EST database as a contig 615 bases in length, and

Table 2.4 *Salmo salar* ESTs retrieved from the cGRASP EST database that were associated with four microarray-identified *γM-crystallin* features (more highly expressed in 50% hatch stage embryos relative to eye-up stage embryos) and used in contig assembly.

***S. salar γM-crystallin* Gene (accession #; contig #)**	**Contig Length (bp)**	**cGRASP cDNA Library Identifier[1]**	**cDNA Library Source**	**Tissue Used for cDNA Library Construction**	**No. of Clones[2]**
***S. salar* cryGM2-like 1 (CB509945; cons47274)**	830	ssal_nwh	Koop Lab	whole juvenile	1
		ssal_hoyc	Hoyheim	eye	2
		ssal_hoyd	Hoyheim	eye	1
		ssal_hoye	Hoyheim	eye	3
		ssal_hoyf	Hoyheim	eye	1
***S. salar* cryGM2-like 2 (CB510589; cons68532)**	629	ssal_nwh	Koop Lab	whole juvenile	1
		ssal_hoyd	Hoyheim	eye	1
		ssal_hoye	Hoyheim	eye	1
		ssal_hoyf	Hoyheim	eye	2
***S. salar* cryGM4-like (CB510792; cons176022)**	615	ssal_nwh	Koop Lab	whole juvenile	1
		ssal_hoyc	Hoyheim	eye	1
		ssal_hoyd	Hoyheim	eye	3
		ssal_hoye	Hoyheim	eye	4
		ssal_hoyf	Hoyheim	eye	1
***S. salar* cryGM3-like (CB511371; cons14521)**	579	ssal_nwh	Koop Lab	whole juvenile	2

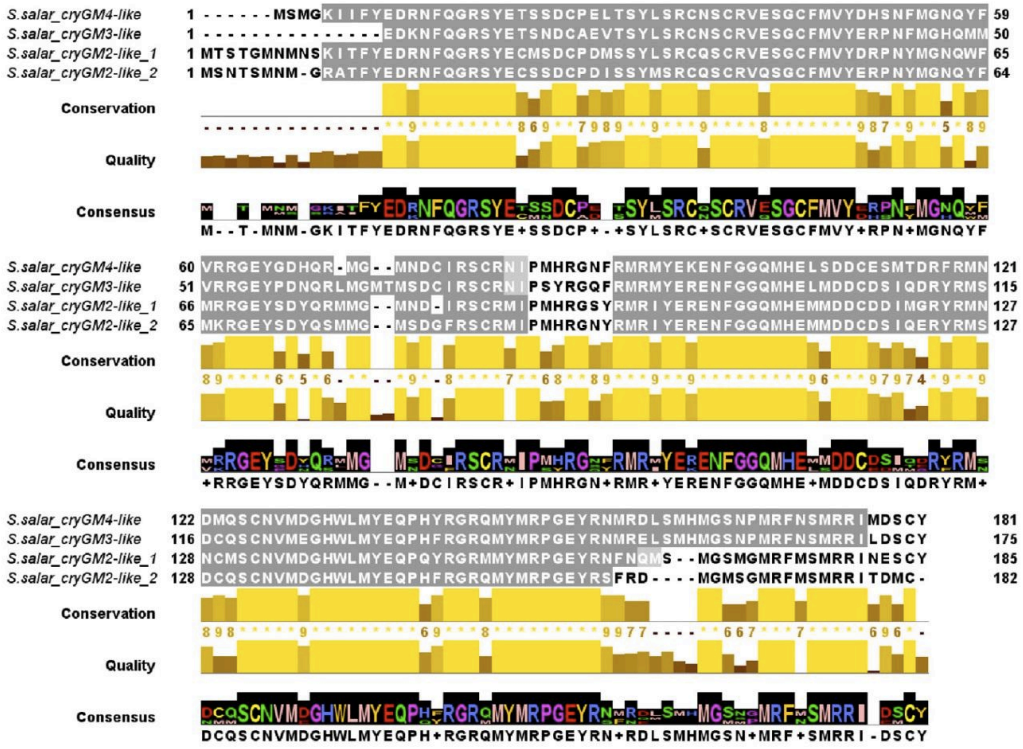
[1]The identifiers (names) of the cDNA libraries in the cGRASP EST database (Web Reference 2.3) are provided in Koop *et al*., 2008.

[2]Some clones were sequenced more than once (forward and reverse), so that the number of clones is not necessarily equal to the number of EST sequences.

named *S. salar cryGM4-like* (CB510792; cons176022). Finally, contig *S. salar cryGM3-like* (CB511371; cons14521) was 579 bases in length, and was assembled from two clones (see Table 2.4). The position of the start codon indicated that while *S. salar* cryGM2-like 1 and 2 are likely complete translated sequences, *S. salar* cryGM3-like sequence is a partial translation (missing the start codon as well as part of the N-terminus encoding region) (Figure 2.4 A). The *S. salar* cryGM4-like sequence starts with a Methionine and, therefore, could include the complete N-terminus (Figure 2.4 A). The Atlantic salmon γM-crystallin paralogues include a highly conserved γ-crystallin domain (Figure 2.4 A). Pairwise alignment of the four protein sequences that represented microarray-identified paralogous γM-crystallin contigs (Figure 2.4 B) showed a close relationship between pairs of paralogues. The complete amino acid sequences of *S. salar* cryGM2-like 1 and *S. salar* cryGM2-like 2 were 79% identical, while the partial amino acid sequences of *S. salar* cryGM3-like and *S. salar* cryGM4-like were 81% identical to one another; all other pairwise alignments returned 71-75% identity (Figure 2.4 B). Phylogenetic tree analysis suggested that there were two sets of more closely related paralogues (Figure 2.4 C), and also indicated that these sequences were more similar to their closest paralogue than to their closest related zebrafish orthologue (e.g., *Salmo salar cryGM2-like 1* and *2* are more similar to one another than either are to *Danio rerio cryGM2c* or *Danio rerio cryGM2f*) (Figure 2.4C).

Figure 2.4 **A)** Predicted amino acid sequences and γ-crystallin domains (in grey; light grey indicates discrepancy in domain position) encoded by the four Atlantic salmon *γM-crystallin* paralogues. Conservation of the consensus sequence is shown in the top histogram and numerical scoring, with an "*" indicating perfect agreement between all sequences. Insertions/deletions (indels) are indicated as "-". The middle histogram shows alignment quality based on Blosum62 scores, with degree of confidence indicated by histogram bar height and color intensity. The consensus sequence is shown by the final histogram, where "+" indicates a tie in the consensus. Residue conservation is indicated by histogram bar height and color intensity. The MSA was created using MUSCLE, and the graphic was created in Jalview 2.6.1. **B)** Percent identity (highest indicated by bold font) between hypothetical amino acid sequences corresponding to four paralogous Atlantic salmon *γM-crystallin* partial or complete cDNA contigs, determined using the NCBI BLAST Align/Blast2seq tool. **C)** Phylogenetic tree (generated by DNASTAR Lasergene 9) showing the evolutionary relationship between the four Atlantic salmon γM-crystallin sequences and the most closely related (Table 2.3) *Danio rerio* γM-crystallin sequences. Values are expressed as percentages following bootstrapping 1000x (with a random seed of 111). A node labeled with an "NA" is a by-product of creating a rooted tree using an un-rooted alignment.

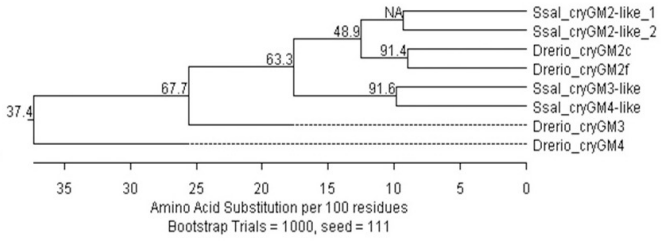
A



B

Aligned *Salmo salar* γM-crystallins	Identities
M2-like_1 and M2-like_2	**79%**
M2-like_1 and M3-like	75%
M2-like_1 and M4-like	71%
M2-like_2 and M3-like	72%
M2-like_2 and M4-like	74%
M3-like and M4-like	**81%**

C



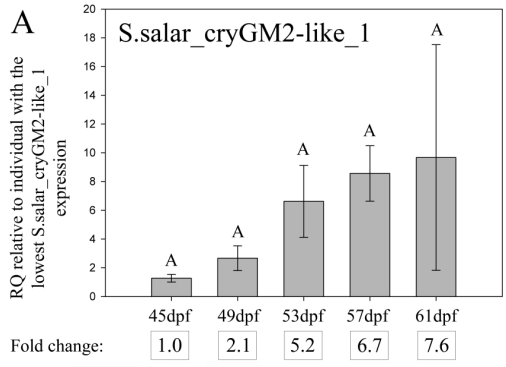
2.4.4 Transcript expression of four putative *γM-crystallin* paralogues during development

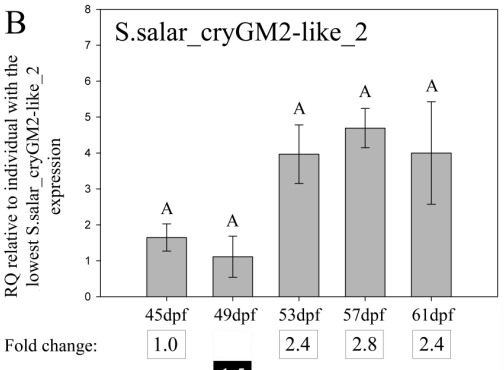
QPCR analysis of the Atlantic salmon *γM-crystallin* paralogues revealed that developmental stage did not have a significant effect on the gene expression of *S. salar cryGM2-like 1* or *cryGM2-like 2* (which share 79% identity at the hypothetical amino acid level) (Figure 2.5 A, B, Table A.1). Transcript expression of *S. salar cryGM4-like* and *S. salar cryGM3-like* (81% identity at the hypothetical amino acid level) increased from 45 dpf to 61 dpf (Figure 2.5 C, D). Peak expression of *S. salar cryGM4-like* occurred at 61 dpf, with a fold change of 18.9 compared to 45 dpf, and was also more highly expressed as compared to 49 dpf embryos ($p < 0.05$) (Figure 2.5 C, Table A.1). Expression of *S. salar cryGM3-like* transcript also increased steadily between 45 and 61 dpf. Peak expression of this transcript also occurred at 61 dpf, with a fold change of 21.6 as compared to at 45 dpf, and expression at 61 dpf was significantly ($p < 0.05$) higher as compared to both 45 dpf and 49 dpf (Figure 2.5 D, Table A.1).

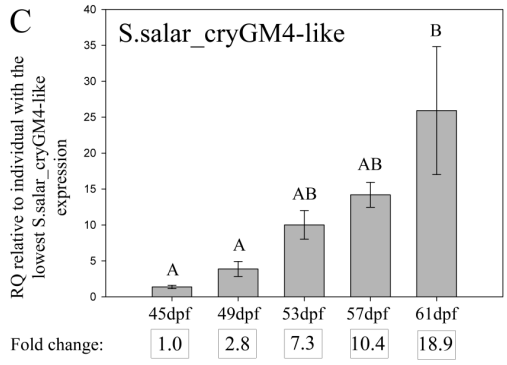
2.4.5 Characterization of four paralogous Atlantic salmon *α-sHSP* contiguous sequences

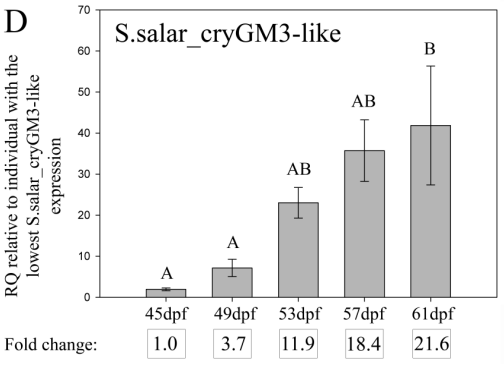
Previous unpublished research on four paralogous rainbow trout *α-sHSP* transcripts (see Table A.2) indicated that these transcripts were synchronously up-regulated just before the hatch event during normal development (Rise and Devlin,

Figure 2.5 QPCR assessment of constitutive expression of four Atlantic salmon *γM-crystallin* paralogues in whole embryos (45 dpf-57 dpf) and early sac fry (61 dpf). Data are presented as mean (± SEM) relative quantity (RQ) normalized to *PABP1* (n = 6: two individuals from three replicate tanks per time point). RQ values were calibrated to the individual with the lowest normalized expression for each target. Bars without a letter in common are significantly different ($p < 0.05$). Overall fold change was determined relative to the initial sampling time point at 45 dpf; fold change in a white box with black text indicates higher expression (although not always statistically significant) at a time point relative to 45 dpf, while fold change in a black box with white text indicates higher expression (although not statistically significant) at 45 dpf relative to 49 dpf. Salmon were eye-up stage at 45 dpf, 50% hatch at 57 dpf, and early sac fry stage at 61 dpf.


A
S.salar_cryGM2-like_1
RQ relative to individual with the lowest S.salar_cryGM2-like_1 expression
20
18
16
14
12
10
8
6
4
2
0
A
A
A
A
A
45dpf
49dpf
53dpf
57dpf
61dpf
Fold change:
1.0
2.1
5.2
6.7
7.6


B
S.salar_cryGM2-like_2
RQ relative to individual with the lowest S.salar_cryGM2-like_2 expression
8
7
6
5
4
3
2
1
0
A
A
A
A
A
45dpf
49dpf
53dpf
57dpf
61dpf
Fold change:
1.0
1.5
2.4
2.8
2.4


C
S.salar_cryGM4-like
RQ relative to individual with the lowest S.salar_cryGM4-like expression
40
35
30
25
20
15
10
5
0
A
A
AB
AB
B
45dpf
49dpf
53dpf
57dpf
61dpf
Fold change:
1.0
2.8
7.3
10.4
18.9


D
S.salar_cryGM3-like
RQ relative to individual with the lowest S.salar_cryGM3-like expression
70
60
50
40
30
20
10
0
A
A
AB
AB
B
45dpf
49dpf
53dpf
57dpf
61dpf
Fold change:
1.0
3.7
11.9
18.4
21.6

unpublished). The cGRASP EST database was mined to identify all currently available *Oncorhynchus mykiss α-sHSP-like* contigs and their contributing ESTs. These ESTs were then assembled into four contigs to confirm the previous *O. mykiss* results (Table A1). These four rainbow trout *α-sHSP-like* contigs were used to identify homologous Atlantic salmon *α-sHSP-like* ESTs (see section 2.3.7 for details). The collected Atlantic salmon sequences were assembled into four paralogous contigs, which were selected *a priori* as genes of interest in this study, and fold-coverage was determined by the total number of clones contributing to a particular contig (Table 2.5).

Alignment of the Atlantic salmon *α-sHSP* paralogous contigs against the nr database using BLASTx (Web Reference 2.1) identified various α-crystallin-domain-containing heat shock proteins, which were functionally annotated with the GO term "response to stress" (Table 2.6). MUSCLE MSA (Web Reference 2.7) of the four Atlantic salmon and rainbow trout α-sHSP contig hypothetical amino acid contig sequences revealed areas of similarity and dissimilarity between the sequences and showed the entire assembled contig and translated amino acid sequences (Figure 2.6 A). Each translation encodes the complete contig (from the start methionine "M" codon, up to, but not including, the stop codon) and a highly conserved α-crystallin domain. The characteristic α-crystallin domain, which is conserved among all α-sHSPs (Caspers *et al.* 1995), was identified in all four paralogous Atlantic salmon sequences, as well as in the associated putative rainbow trout orthologues (Figure 2.6

Table 2.5 Four paralogous *Oncorhynchus mykiss α-sHSP* contigs, and the cDNA libraries that contributed ESTs from the cGRASP EST database for putative orthologous *Salmo salar α-sHSP* contigs.

O. mykiss α-sHSP **Gene (contig #)[1]**	***S. salar α-sHSP*** **Gene (contig #)**	**Contig Length (bp)**	**cGRASP cDNA Library Identifier[2]**	***S. salar*** **Tissue Used to Generate cDNA Library**	**No. of Clones[3]**
O.mykiss α-sHSP 1 (cons26785)	***S.salar* α-sHSP 1 (cons33468)**	956	ssal_evd	thymus	2
			ssal_hrmf	red muscle	1
O.mykiss α-sHSP 2 (cons30611)	***S.salar* α-sHSP 2 (cons69807)**	1001	ssal_evd	thymus	4
			ssal_nwh	whole juvenile	1
O.mykiss α-sHSP 3 (cons111289)	***S.salar* α-sHSP 3 (cons14647)**	895	ssal_can	cardiac	3
			ssal_hhud	skin	1
			ssal_evd	thymus	4
			ssal_eve	thyroid	6
O.mykiss α-sHSP 4 (cons21483)	***S.salar* α-sHSP 4 (cons81495)**	936	ssal_evd	thymus	5
			ssal_eve	thyroid	2
			ssal_hhe	heart	2
			ssal_plnb	pyloric caeca	1

[1]Contig numbers (cons#) were automatically assigned by the cGRASP EST database (Web Reference 2.3).

[2]The identifiers (names) of the cDNA libraries in the cGRASP EST database (Web Reference 2.3) were provided in Koop *et al*. (2008).

[3]Some clones were sequenced more than once (forward and reverse), so that the number of clones is not necessarily equal to the number of EST sequences.

Table 2.6 Best BLASTx hit[1] and functional annotation of four paralogous Atlantic salmon *α-sHSP* transcripts.

Gene (contig number)	**Best BLASTx Hit[1] (species)**	**BLASTx Hit Accession Number**	**Percent Identity (aligned residues)**	**E-value**	**Function of Best BLASTx Hit[2]**	**GO Identifier**
***S.salar* α-sHSP 1 (cons33468)[3]**	Heat shock protein 30 (*Salmo salar*)	ACM09782	99% (239/240)	1e-118	BP: Response to stress	GO:0006950
***S.salar* α-sHSP 2 (cons69807)**	Heat shock protein Hsp-16.1/ Hsp-16.11 (*Salmo salar*)	NP_001134496	100% (235/235)	1e-117	BP: Response to stress	GO:0006950
***S.salar* α-sHSP 3 (cons14647)**	Heat shock protein beta-11 (*Salmo salar*)	ACI69564	100% (206/206)	8e-104	BP: Response to stress	GO:0006950
***S.salar* α-sHSP 4 (cons81495)**	Heat-shock protein 30-like (*Salmo salar*)	ACH70863	100% (209/209)	2e-118	BP: Response to stress	GO:0006950

[1]Best BLASTx hit defined as the best (lowest E-value) hit with an associated protein or gene name (not "predicted" or "hypothetical") with an E-value of 1e-10 or less. BLASTx reports were collected on May 27th, 2011, and reflect the entries collected in the GenBank non-redundant (nr) amino acid database to this date.

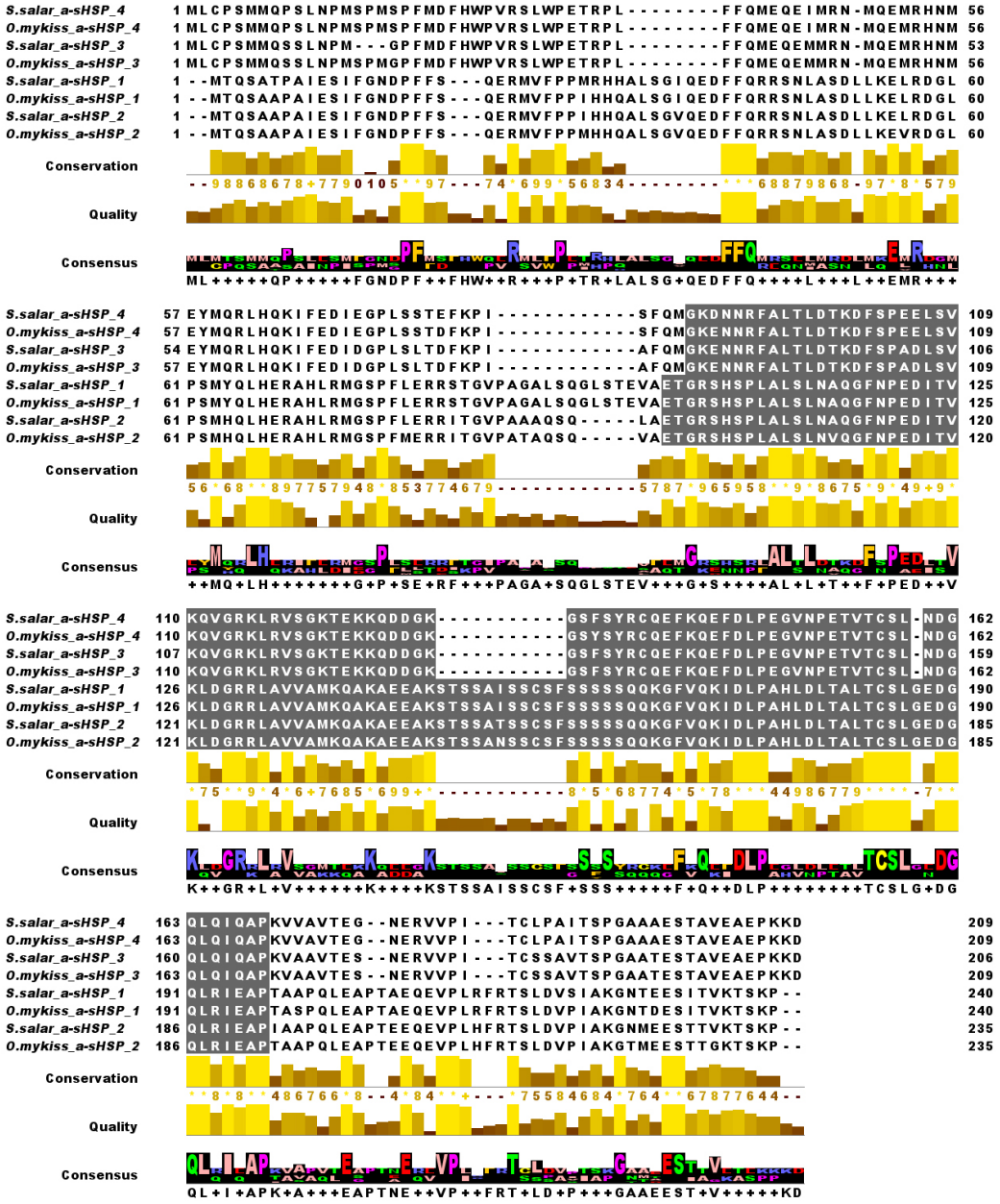
[2]The functional annotations were obtained from the UniProt Knowledgebase (Web Reference 2.9). Gene ontology (GO) terms (biological process (BP)) are associated with a corresponding GO identifier.

[3]The assigned name (e.g., *S. salar α-sHSP 1*) and the contiguous sequence (contig) number (e.g., contig consensus sequence 33468) of each gene is given. Contributing ESTs were retrieved (see Table A1) from the cGRASP EST database (Web Reference 2.3), which performed automated sequence assembly of the associated ESTs into contigs.

Figure 2.6 **A)** Predicted amino acid sequences and α-crystallin domains (in grey) encoded by four rainbow trout and four Atlantic salmon α-sHSP paralogues. Domain position was obtained from BLASTp and the UniProtKb. Conservation (indicated by bar height and color intensity) of the consensus sequence is shown in the top histogram and is also reflected by a numerical scoring, with an "*" indicating perfect agreement between all sequences. Indels are indicated as "-". Quality, the middle histogram, shows alignment quality based on Blosum62 scores, with degree of confidence indicated by histogram bar height and color intensity. The consensus sequence is shown at the bottom, accompanied by a final histogram, where "+" indicates a tie in the consensus. The sequences were aligned with MUSCLE (Web Reference 2.7), and the graphic was created in Jalview 2.6.1. **B)** Percent identity between hypothetical amino acid sequences corresponding to four complete (based on alignment with putative orthologous *Danio rerio* complete cDNAs) paralogous Atlantic salmon α-sHSP sequences, determined using the NCBI BLAST Align/ Blast2seq tool (Web Reference 2.1). The highest percent identity between pairs of paralogues is indicated by a bold font. **C)** Phylogenetic tree showing the evolutionary relationship between four Atlantic salmon and four rainbow trout paralogous α-sHSP proteins (as determined by alignment with putative orthologous *Danio rerio* complete protein sequences found in GenBank). Values are expressed as percentages following bootstrapping 1000 times (with a random seed of 111) to assess confidence in the construction of the tree. A node labeled "NA" is a by-product of creating a

rooted tree using the un-rooted alignment created by MUSCLE. The tree was generated by DNASTAR Lasergene 9 software.

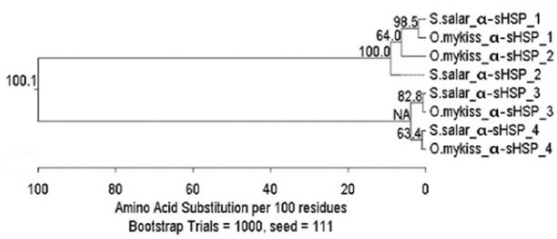
A



B

Aligned *Salmo salar* α-sHSPs	Identities
1 and 2	**81%**
1 and 3	32%
1 and 4	28%
2 and 3	31%
2 and 4	26%
3 and 4	**83%**

C

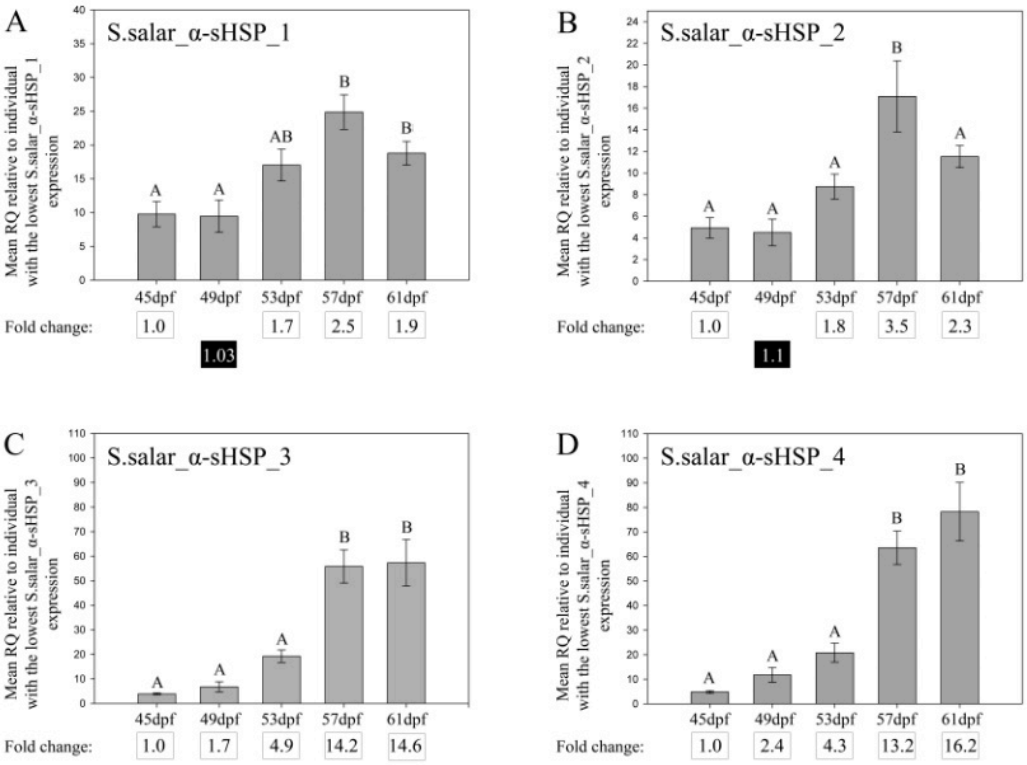


A). The MSA found that structural similarity was highest between *S. salar* paralogues α-sHSP 1 and 2 and between *S. salar* α-sHSP 3 and 4, at both the nucleotide (93 and 92%, respectively) and putative protein (81 and 83%, respectively) levels. Other pairwise alignments (e.g., between *S. salar* α-sHSP 1 and 3) showed much lower levels of conservation (26-32% at the hypothetical amino acid level) (Figure 2.6 B). The molecular phylogenetic tree confirmed that, while Atlantic salmon and rainbow trout both had pairs of relatively closely related paralogues, the closest relationships were between corresponding putative orthologous amino acid sequences from the two species (Figure 2.6 C).

2.4.6 Transcript expression of four *α-sHSP* paralogues during developmental stages surrounding the hatch event

Constitutive transcript expression for all four paralogues was studied at five developmental time points from 45 dpf (eye-up stage) to 61 dpf (post-hatch) (n = 6 individuals per time point for each target) (Figure 2.7). QPCR analysis of the Atlantic salmon *α-sHSP* paralogues revealed that developmental stage had a significant effect on the gene expression of all four paralogues (Table A.1). Significant, and synchronous, up-regulation of all four transcripts occurred at 57 dpf (50% hatch) compared to 45 dpf (eye-up stage) (Figure 2.7, Table A.1). Expression profiles and the magnitude of up-regulation were most similar for pairs of paralogues that were more closely related. *S. salar α-sHSPs 1* and *2* share 81% identity at the hypothetical amino acid level (Figure 2.6 B), and have similar developmental transcript

Figure 2.7 QPCR assessment of constitutive transcript expression of four Atlantic salmon *α-sHSP* paralogues in whole embryos (45 dpf to 57 dpf) and sac fry (61 dpf). Data are presented as mean (± SEM) relative quantity (RQ) normalized to *PABP1* (n = 6: two individuals from three replicate tanks per time point). RQ values were calibrated to the individual with the lowest normalized expression for each target. Bars without a letter in common are significantly different ($p < 0.05$). Fold change was determined relative to the initial sampling time point (45 dpf). Fold change in a white box with black text indicates higher expression (although not always significant) at a time point compared to 45 dpf, while fold change in a black box with white text indicates lower expression (not significant) at a time point compared to 45 dpf. Fish were at eye-up stage at 45 dpf, 50% hatch stage at 57 dpf, and early sac fry stage at 61 dpf.


A
S.salar_α-sHSP_1
Mean RQ relative to individual with the lowest S.salar_α-sHSP_1 expression
A A AB B B
45dpf 49dpf 53dpf 57dpf 61dpf
Fold change: 1.0 1.03 1.7 2.5 1.9
B
S.salar_α-sHSP_2
Mean RQ relative to individual with the lowest S.salar_α-sHSP_2 expression
A A A B A
45dpf 49dpf 53dpf 57dpf 61dpf
Fold change: 1.0 1.1 1.8 3.5 2.3
C
S.salar_α-sHSP_3
Mean RQ relative to individual with the lowest S.salar_α-sHSP_3 expression
A A A B B
45dpf 49dpf 53dpf 57dpf 61dpf
Fold change: 1.0 1.7 4.9 14.2 14.6
D
S.salar_α-sHSP_4
Mean RQ relative to individual with the lowest S.salar_α-sHSP_4 expression
A A A B B
45dpf 49dpf 53dpf 57dpf 61dpf
Fold change: 1.0 2.4 4.3 13.2 16.2

expression profiles: *S. salar α-sHSP 1* and *α-sHSP 2* transcript expression peaked at 57 dpf (overall 2.5-fold and 3.5-fold, respectively, higher expression compared with 45 dpf) (Figure 2.7 A, B). *S. salar α- sHSP 1* transcript expression was also higher at 57 dpf compared to 49 dpf embryos (Figure 2.7 A), whereas *S. salar α-sHSP 2* transcript expression was significantly higher at 57 dpf compared with 49 and 53 dpf (Figure 2.7 B). *S. salar α-sHSP 3* and *α-sHSP 4* share 83% identity at the hypothetical amino acid level (Figure 2.6 B). Peak expression occurred at 61 dpf (overall 14.6-fold and 16.2-fold, respectively, more highly expressed compared with 45 dpf), with significantly higher expression also at both 57 dpf and 61 dpf stages as compared to 49 and 53 dpf (Figure 2.7 C, D).

2.5 DISCUSSION

2.5.1 *γM crystallin* paralogues: evidence of gene duplication and divergence, molecular phylogenetics, and functional significance

As a large family of radiated genes, many of which currently have unknown functions, the family of γ-crystallins can be used to provide information on gene duplication and divergence. To this end, microarray-identified *γM crystallin* transcripts were selected for paralogue discovery and transcript expression analyses.

The evidence of gene duplication and divergence of *γM crystallin* paralogues presented in this study is threefold. Firstly, we examined structural similarity among

sequences and identified highly conserved regions characteristic of the γ-crystallin two-domain structure; each domain is composed of two Greek key motifs, and the domains are separated by a short polypeptide (Slingsby *et al.* 2013). In agreement with a study by Koop *et al.* (2008), which found that 23,892 of the 81,398 Atlantic salmon contigs studied shared 80-96.9% identity with at least one other contig, this study found that the pairs of more similar γM-crystallins shared ~80% identity at the predicted amino acid level; more dissimilar paralogues were 71-74% similar. Secondly, pairs of more highly similar paralogues shared more similar expression and fold-induction profiles surrounding the hatch event in Atlantic salmon. For example, the maximum fold induction of *cryGM2-like 1* (7.6-fold) and *cryGM2-like 2* (2.8-fold) was lower (and non-significant) as compared to the significantly higher expressed maximum fold induction of *cryGM3-like* (21.6-fold), and *cryGM4-like* (18.9-fold) (Figure 2.7). These findings are a strong indication that the features identified in this study represent paralogous transcripts and the structural similarities and differences among the *γM-crystallins* examined in this study are echoed in their respective expression profiles. Thirdly, phylogenetic relationships generated in this study suggested that pairs of more highly similar Atlantic salmon γM-crystallin paralogues were more closely related to each other than to their closest related zebrafish putative orthologues. However, confidence in the phylogenetic tree was limited due to some low bootstrap values (e.g., 48.9) that indicated branch points of lower trust. These findings suggest that some *γM-crystallin* gene duplication events

(e.g. giving rise to GM2-like 1 and GM2-like 2) occurred within the salmonid lineage (i.e., after the divergence of Atlantic salmon and zebrafish), while the divergence between *γM2-like crystallins* and *γM3* and *4-like crystallins* likely occurred before speciation.

As a highly radiated gene family, examining the relationships among paralogous γ-crystallins provides evidence of gene duplication and divergence as both a measure and a means of evolution. For example, previous studies have examined members of the γ-crystallin family in order to resolve evolutionary relationships (Chiou 1986; Chiou *et al.* 1986b; Kappé *et al.* 2010; McFall-Ngai and Horwitz 1990; Slingsby *et al.* 2013). Multiple γ-crystallin paralogues have been identified in the carp (*Cyprinus carpio*) (Chiou *et al.* 1986b; Chang *et al.* 1991), and have been shown to be closely related to orthologous γ-crystallins in haddock (*Melanogrammus aeglefinus*), squid (*Loligo pealii*) (Chiou *et al.* 1986a), calf (*Bos taurus*), frog (*Xenopus laevis)* (Chiou *et al.* 1986b) and shark (*Chiloscyllium colax*) (Chiou *et al.* 1990). These findings suggest a common ancestral γ-crystallin early in the bilateria lineage, with subsequent divergence in the evolution of this gene family (Chiou 1986; Chiou *et al.* 1986a, b; Chiou 1990; McFall-Ngai and Horwitz 1990; Kappé *et al.* 2010; Slingsby *et al.* 2013). Each *γM-crystallin* paralogue in this study represents a member of the aquatic M family, one of the four classes of γ-crystallins (the aquatic γM class; the terrestrial class of γ-crystallins such as the mammalian γA-F class; the γS class; and the γN class) (Wistow *et al.* 2005). Collectively, the

structural, phylogenetic, and expression data obtained in the current study provide evidence of divergence (e.g., in structure, gene regulation and, potentially, in function) after gene duplication.

Understanding the functions of genes that are up-regulated at the time of hatching (and potentially involved in dealing with the stress of the hatch event) could impart important information about the evolution of fitness-associated transcripts at critical life stages, particularly in highly radiated gene families such as the γM-crystallins. Although functional annotations were obtained for the majority of the microarray-identified features, no functional annotation was available for any of the eight *γM-crystallin-like* transcripts that were identified as more highly expressed at 50% hatch compared to the eye-up stage. Their putative orthologues are known to function in lens formation in vertebrates, such as humans and zebrafish (Greiling *et al.* 2009; Wistow *et al.* 2005). For example, a study by Wages *et al.* (2013) used two-dimensional gel electrophoresis and mass spectrometry to study the presence of crystallin proteins in the lens of zebrafish as they age. Their study revealed three groups of crystallins: the first group of crystallins was expressed consistently from 4.5 days old to 27 months old; the second group was expressed before 10 weeks old; the third group was expressed after 10 weeks old. In this study, the aquatic γM-crystallins belonged to the second temporal grouping of proteins and were already abundant at the initial 45 dpf time point. In addition to determining that different crystallins are present at different developmental stages, Greiling *et al.* (2009) used

mass spectrometry to determine that the presence of γ-crystallins in the zebrafish lens increased with age. Two-dimensional gel electrophoresis found that γ-crystallins of adult zebrafish comprised 42% of lens proteins, and 30.5% were specifically γM-crystallins (Posner *et al.* 2008); this high abundance may allow dense packing of the lens proteins and enable underwater vision (Kiss *et al.* 2004; Wistow *et al*. 2005). Structurally, the γM-crystallins of the fish lens can be distinguished from other vertebrate γ-crystallins by the partial loss of four conserved mammalian tryptophan residues that are believed to protect the lens from UV light (Mahler *et al.* 2013; Pan *et al.* 1994). The chemical properties of γ-crystallin proteins include high solubility and thermodynamic stability, characteristics that permit transparency and the high light refractive index needed for sight (Zhao *et al.* 2011). The *γM-crystallins* in this study have similar structural characteristics and are potentially involved in similar functions (i.e., lens formation, ontogeny of vision)(Easter and Nicola 1996).

Several published studies show that the disruption of the constitutive expression of the γ-crystallins can be detrimental to their roles in lens development and vision. For example, γ-Crystallin protein content was found to be diminished in the zebrafish *cloche* mutant, which displays cataract formation and increased reflection of the lens (Goishi *et al.* 2006). Crystallin defects have also been identified as a contributing factor in cataract formation in mouse and human, with at least 20 mutations identified in mouse and mutations in *γC-crystallin* and *γD-crystallin* identified in humans (Graw 2004). Collectively, these studies show that ontological

expression of crystallins in the zebrafish eye is highly regulated, and that the disruption of ontological expression of genes potentially involved in lens formation and vision would potentially have an impact on fitness. Despite the studies mentioned above, little is known about the function and expression of γ-crystallins during embryonic vertebrate development, and even less is known about the functional differences between paralogous members of this expanded gene family. Further research is needed to determine the functional significance of fish γM-crystallins, and of the distinct developmental expression profiles of the salmon *γM-crystallin* paralogues observed in this study.

2.5.2 *α-sHSP* paralogues: evidence of gene duplication and divergence, molecular phylogenetics, and functional significance

As with the γM-crystallins, structural (sequence analysis), gene expression, and molecular phylogenetics provided potential evidence of gene duplication and divergence of Atlantic salmon α-sHSPs. Each α-sHSPs contained an α-crystallin domain, which is conserved amongst all α-sHSPs (Caspers *et al.* 1995; Narberhaus 2002). While the four Atlantic salmon *α-sHSP* transcripts were up-regulated at hatch compared with earlier developmental stages, developmental transcript expression profiles were most similar among pairs of paralogues with the highest structural similarity. Phylogenetic analysis suggested that an ancient gene duplication event (possibly due to the fish-specific WGD event) led to the existence of two *α-sHSP*

gene copies, and that subsequent gene duplication events (possibly due to the salmonid-specific WGD event), before the divergence of genera *Salmo* and *Oncorhynchus*, led to the existence of these four *α-sHSP* gene copies in Atlantic salmon and rainbow trout genomes.

The radiation of the sHSP family has been studied previously in the tetrapod lineage, in zebrafish and other teleost fish, and in other vertebrates and provides evidence of gene duplication and divergence (Reviewed in Basu *et al.* 2002; Heikkila *et al.* 1986; Narberhaus 2002; Narum and Campbell 2010). For example, 10 sHSPs have been identified in humans, whereas 13 sHSP paralogues have been identified in zebrafish (Elicker and Hutson 2007). Although zebrafish *sHSP* transcripts with orthologous human counterparts have been discovered, such as *HSPB13* (alias *HSPB6*) and *HSPB14* (alias *HSPB9*), some teleost *sHSPs* (e.g., fish *HSPB15*) have no known tetrapod orthologues, and some tetrapod *sHSPs* (e.g., mammalian *HSPB10*) have no known teleost fish orthologues (Elicker and Hutson 2007; Franck *et al.* 2004; Huang *et al.* 2008). Of particular relevance, the *hsp30*-like transcripts (alias *hspb11* or *hsp25*) examined in this thesis have no mammalian counterparts, but are present in non-mammalian vertebrates (reviewed in Heikkila 2017).

The measure of structural (i.e., sequence) similarity between homologous genes is an indication of evolutionary distance, as more mutations occur over evolutionary time. Therefore, more similar paralogues have likely arisen from more recent gene duplication events (Koop *et al.* 2008). For example, the overall similarity

between Atlantic salmon and rainbow trout (two closely related species) orthologous ESTs is approximately 93% (Koop *et al.* 2008). A study by Garcia de la Serrana and Johnston (2013) reported similarity among six Atlantic salmon HSP90 paralogues: four HSP90α amino acid sequences shared 87-97% identity, while two HSP90β amino acid sequences shared 98% identity. However, alignments between HSP90α and β amino acid sequences revealed 82-84% identity between these groups of paralogues (Garcia de la Serrana and Johnston 2013). Each of the six Atlantic salmon HSP90 paralogues were more closely related to orthologous rainbow trout counterparts than to each other, as determined by phylogenetic analysis (Garcia de la Serrana and Johnston 2013). In agreement with the studies described above, the results of the current study also provide further information on the evolutionary relationships between homologous members of the sHSP family, which in turn supports our understanding of gene duplication and speciation events in the salmonid lineage. Additionally, the expression profiles of these genes during late embryonic development, the hatch event, and in post-hatch Atlantic salmon provide information on ontogenic expression and potential function of these transcripts.

HSPs have been previously studied during normal (i.e., unstressed) development in zebrafish and other vertebrates (Elicker and Hutson 2007; reviewed in Heikkila 2017; Morimoto *et al.* 1990; reviewed in Rupik *et al.* 2011). Previous studies found that *HSP* expression is both life-stage and species-dependant (e.g., see Heikkila *et al.* 1986). In zebrafish, for example, constitutive expression of *hsp27* was

induced during early gastrulation and was expressed in embryos following heat stress (Mao and Shelden 2006), while constitutive and heat stress-induced expression of *Xenopus hsp27* was first observed at tail bud stage (Tuttle *et al.* 2007). Each of the *α-sHSPs* in the current study were expressed at the eye-up stage (45 dpf) and encoded a conserved α-crystallin domain; in humans, α-A-crystallin (*HSPB4*) and α-B-crystallin (*HSPB5*) have been linked to lens formation of the eye (Narberhaus 2002), and mutations in these transcripts are linked with cataract development, as was previously discussed (Berry *et al.* 2001; Litt *et al.* 1998). The results of the current study provide novel information specific to the constitutive transcript expression dynamics and functions of *α-sHSPs* during early development in Atlantic salmon, and these transcripts, like the *γM-crystallins*, may also function in sight.

Although all four Atlantic salmon *α-sHSPs* were expressed under normal hatching conditions in Atlantic salmon (current study) and in rainbow trout (Rise and Devlin, unpublished), these genes were functionally annotated with stress response identifiers. α-sHSPs play roles in anti-apoptosis, the negative regulation of intracellular transport, protein folding, protein binding and unfolded protein binding (UniProtKb; Web Reference 2.9). Further, they are well known for their activity as molecular chaperones during heat stress, where they prevent incorrect tertiary folding of denatured proteins (Heikkila *et al.* 1986; Narum and Campbell 2010). The molecular chaperone function of the HSP gene family has also been studied in response to hypoxia, exposure to toxins, and infection (reviewed in Basu *et al.* 2002;

Elicker and Hutson 2007; reviewed in Heikkila *et al.* 1986; Mao *et al.* 2005). It has been postulated that the rainbow trout putative orthologues of the Atlantic salmon *α-sHSPs* examined in this study accumulate before hatch as a pre-adaptation to this potentially stressful event, which involves the embryos leaving the protection of the egg and potentially entering a challenging environment (Rise and Devlin, unpublished). The expression of the suite of four Atlantic salmon paralogues was also significantly higher at hatch event compared to earlier developmental stages, supporting this hypothesis.

The current *α-sHSP* transcript expression results appear to provide evidence both of conserved regulation among *α-sHSP* paralogues (e.g., the very similar developmental transcript expression profiles of *α-sHSP 1* and *2*) as well as potential regulatory divergence of some *α-sHSP* paralogues (e.g., average fold-induction during hatch of 3.5 and 13.2 and different timing of peak expression for *α-sHSP 2* and *4*, respectively). The divergence in the expression profiles among paralogous transcripts provides some evidence of neofunctionalization. Alternatively, the increase in the expression of this suite of paralogues surrounding the hatch event could suggest that they retain some ancestral functions. The expression profile of these genes may also suggest that in the presence of multiple paralogues and decreased conservation pressure, the known stress response function of these transcripts has been re-appropriated to function as a pre-adaptation to hatch stress.

2.5.3 Other evidence of gene duplication and divergence in fish

There is ample evidence supporting the occurrence of the FSGD event 320-350 m.y.a. in the teleost lineage and the salmonid-specific WGD event 80 m.y.a., the latter explaining the pseudotetraploid nature of salmonid genomes (Allendorf and Thorgaard 1984; Christoffels *et al.* 2004; Davidson *et al.* 2010; Lien *et al.* 2016; Meyer and Van de Peer 2005; Ohno 1970; Schartl 1999; Steinke *et al.* 2006a, b; Taylor *et al.* 2003). For example, in a study by Jiang *et al.* (2015), molecular phylogenetic analysis of interleukin-12 alpha chain sequences showed evidence of gene or genome duplication common to all teleosts (potentially arising from the FSGD) as well as gene or genome duplications within the salmonid lineage.

Previous studies involving fish as models have used transcript expression of paralogues to provide evidence of gene duplication and divergence (potentially pointing to neofunctionalization). For example, Xue *et al*. (2014) used RT-PCR with templates from 15 different tissues to show that Atlantic cod (*Gadus morhua*) paralogues *elovl1a* and *elovl1b* have very different constitutive transcript expression across tissues; *elovl1a* had relatively high expression in posterior kidney, stomach and gill and extremely low or absent expression in some tissues (e.g., brain, liver and heart), whereas *elovl1b* had relatively high transcript expression in all 15 tissues. Furthermore, Xue *et al*. (2014) interpreted these differences in basal expression as suggestive of regulatory, and possibly functional, divergence of the paralogues. Marked differences were also seen in the tissue specific transcript expression of 4

cod *elovl4* paralogues (Xue *et al.* 2014), providing further evidence of divergence after gene duplication. Based on differences in liver transcript expression profiles of two beta-2-microglobulin paralogues (B2M1 and B2M2, 91% identical at the hypothetical amino acid level) among families of triploid growth hormone transgenic Atlantic salmon, Xu *et al*. (2013) speculated that these duplicated genes may have diverged functionally. Additionally, Inkpen *et al*. (2015) showed that Atlantic cod IRF4 paralogues (IRF4a and IRF4b, 74% identical at the amino acid level) had very different transcript expression responses to immune stimuli (i.e., only IRF4b was significantly up-regulated by viral or bacterial antigens), and provided evidence of gene duplication and divergence. Finally, studies on the fatty-acid binding protein (fabp) multigene family in zebrafish, rainbow trout, and other teleosts revealed that these highly structurally and functionally conserved genes are differentially regulated. These studies provided further evidence of WGD events and illustrated how subfunctionalization of regulatory factors can influence the fates of duplicated genes (Bayır *et al.* 2015; Laprairie *et al.* 2016a and b; Parmar and Wright 2013; Thirumaran *et al.* 2014). The current study's developmental transcript expression profiles for *γM-crystallins* and *α-sHSPs* provide further evidence of regulatory divergence of paralogues after gene/genome duplication events.

2.6 **CONCLUSIONS**

This study identified 11 transcripts with lower expression and 73 transcripts with higher expression in 50% hatch stage Atlantic salmon embryos relative to eye-

up stage embryos. The differentially expressed genes identified in this study included suites of potential paralogues, including eight *γM-crystallin* –like features, twelve *parvalbumin-like* features, and four features identified as *glyceraldehyde-3-phosphate dehydrogenase.* Analysis of paralogous suites of Atlantic salmon *γM-crystallins* and *α-sHSPs* revealed divergence in expression among suites of paralogues, and also demonstrated that more structurally similar paralogues had more similar expression profiles. The functional genomics research presented in this chapter contributes to the available data on gene expression surrounding the hatch event, and provides insight on paralogue divergence. Further investigation is needed to compare the regulatory regions of these genes, as well as the potential influence of differential expression profiles on fitness.

2.7 REFERENCES

Allendorf, F.W., Thorgaard, G.H. (1984). Tetraploidy and the evolution of salmonid fishes. In: Evolutionary genetics of fishes (Ed. B.J. Turner), Plenum Press, New York, 1-53.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410.

Bailey, G.S., Poulter, R.T., Stockwell, P.A. (1978). Gene duplication in tetraploid fish: model for gene silencing at unlinked duplicated loci. *Proceedings of the National Academy of Sciences USA* 75:5575–5579.

Basu, N., Todgham, A.E., Ackerman, P.A., Bibeau, M.R., Nakano, K., Schulte, P.M., Iwama, G.K. (2002). Heat shock protein genes and their functional significance in fish. *Gene* 295:173–183.

Bayır, M., Bayır, A., Wright, J.M. (2015). Divergent spatial regulation of duplicated fatty acid-binding protein (*fabp*) genes in rainbow trout (*Oncorhynchus mykiss*). *Comparative Biochemistry and Physiology, Part D: Genomics and Proteomics* 14:26-32.

Berry, V., Francis, P., Reddy, M.A., Collyer, D., Vithana, E., MacKay, I., *et al.* (2001). Alpha-B crystallin gene (CRYAB) mutation causes dominant congenital posterior polar cataract in humans. *American Journal of Human Genetics* 69:1141–1145.

Booman, M., Borza, T., Feng, C.Y., Hori, T.S., Higgins, B., Culf, A., *et al.* (2011). Development and experimental validation of a 20K Atlantic cod (*Gadus morhua*) oligonucleotide microarray based on a collection of over 150,000 ESTs. *Marine Biotechnology* 13(4): 733-750.

Bowtell, D., Sambrook, J., Eds. (2003). DNA Microarrays: A Molecular Cloning Manual. Cold Spring Harbour Laboratory Press, New York.

Caspers, G.J., Leunissen, J.A., de Jong, W.W. (1995). The expanding small heat-shock protein family, and structure predictions of the conserved "alpha-crystallin domain". *Journal of Molecular Evolution* 40(3):238-248.

Chang, T., Lin, C.L., Chen, P.H., Chang, W.C. (1991). γ-Crystallin genes in carp: cloning and characterization. *Biochimica et Biophysica Acta* 1090(2):261-264.

Chiou, S.H. (1986). Phylogenetic comparison of lens crystallins from the vertebrate and invertebrate – convergent or divergent evolution? *FEBS Letters* 201(1):89-73.

Chiou, S.H., Chang, T., Chang, W.C., Kuo, J., Lo, T.B. (1986a). Characterization of lens crystallins and their mRNA from the carp lenses. *Biochimica et Biophysica Acta* 871(3):324-328.

Chiou, S.H., Chen, S.W., Itoh, T., Kaji, H., Samejima, T. (1990). Comparison of the γ-crystallins isolated from eye lenses of shark and carp. Unique secondary

and tertiary structure of γ-Crystallin. *FEBS Letters* 275(1-2):111-113.

Chiou, S.H., Chen, S.W., Lo, T.B. (1986b). The amino-terminal sequences of four major carp γ-crystallin polypeptides and their homology with frog and calf γ-crystallins. *FEBS Letters* 209(1):107-110.

Christoffels, A., Koh, E.G.L., Chia, J., Brenner, S., Aparicio, S., Venkatesh, B. (2004). Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Molecular Biology and Evolution* 21(6):1146-1151.

Davidson, W.S., Koop, B.F., Jones, S.J.M., Iturra, P., Vidal, R., Maass, A., *et al.* (2010). Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biology* 11:403-409.

Easter, S.S., Jr., Nicola, G.N. (1996). The development of vision in the zebrafish (*Danio rerio*). *Developmental Biology* 180(2): 646–663.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5):1792-1797.

Elicker, K.S., Hutson, L.D. (2007). Genome-wide analysis and expression profiling of the small heat shock proteins in zebrafish. *Gene* 403(1-2):60-69.

Emerson, K., Russo, R.C., Lund, R.E., Thurston, R.V. (1975). Aqueous ammonia equilibrium calculations: effect of pH and temperature. *Journal of the Fisheries Research Board of Canada* 32(12):2379-2383.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39(4):783-791.

Franck, E., Madsen, O., van Rheede, T., Ricard, G., Huynen, M.A., de Jong, W.W. (2004). Evolutionary diversity of vertebrate small heat shock proteins. *Journal of Molecular Evolution* 59(6):792-805.

Garcia de la Serrana, D., Johnston, I.A. (2013). Expression of heat shock protein (HSP90) paralogues is regulated by amino acids in skeletal muscle of Atlantic salmon. *PLoS One* 8(9):e74295.

Goishi, K., Shimizu, A., Najarro, G., Watanabe, S., Rogers, R., Zon, L.I., Klagsbrun, M. (2006). αA-crystallin expression prevents γ-crystallin insolubility and cataract formation in the zebrafish *cloche* mutant lens.

Development 133: 2585-2593.

Graw, J., Neuhäuser-Klaus, A., Klopp, N., Selby, P.B., Löster, J., Favor, J. (2004). Genetic and allelic heterogeneity of Cryg mutations in eight distinct forms of dominant cataract in the mouse. *Investigative Ophthalmology and Visual Science* 45(4):1202–1213.

Greiling, T.M., Houck, S.A., Clark, J.I. (2009). The zebrafish lens proteome during development and ageing. *Molecular Vision* 15:2313-2325.

Hall, J.R., Clow, K.A., Rise, M.L., Driedzic, W.R. (2011). Identification and validation of differentially expressed transcripts in a hepatocyte model of

cold-induced glycerol production in rainbow smelt (*Osmerus mordax*). *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology* 301(4):R995-R1010.

Heikkila, J.J. (2017). The expression and function of *hsp30*-like small heat shock protein genes in amphibians, birds, fish, and reptiles. *Comparative Biochemistry and Physiology, Part A: Molecular and Integrative Physiology* 203:179-192.

Heikkila, J.J., Browder, L.W., Gedamu, L., Nickells, R.W., Schultz, G.A. (1986). Heat-shock gene expression in animal embryonic systems. *Canadian Journal of Genetics and Cytology* 28(6):1093-105.

Huang, L., Wang, H.S., Kang, L. (2008). Different evolutionary lineages of large and small heat shock proteins in eukaryotes. *Cell Research* 18:1074–1076.

Inkpen, S.M., Hori T.S., Gamperl, A.K., Nash, G.W., Rise, M.L. (2015). Characterization and expression analyses of five interferon regulatory factor transcripts (Irf4a, Irf4b, Irf7, Irf8, Irf10) in Atlantic cod (*Gadus morhua*). *Fish and Shelfish Immunology* 44(1):365-81.

Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N., Mauceli, E., *et al.* (2004). Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431:946-957.

Jiang, Y., Husain, M., Qi, Z., Bird, S., Wang, T. (2015). Identification and expression analysis of two interleukin-23α (p19) isoforms, in rainbow trout *Oncorhynchus mykiss* and Atlantic salmon *Salmo salar*. *Molecular Immunology* 66(2):216-228.

Johnston, I.A., Strugnell, G., McCracken, M.L., Johnstone, R. (1999). Muscle growth and development in normal-sex-ratio and all-female diploid and triploid Atlantic salmon. *Journal of Experimental Biology* 202(15): 1991-2016.

Kappé, G., Purkiss, A.G., van Genesen, S.T., Slingsby, C., Lubsen, N.H. (2010). Explosive expansion of βγ-crystallin genes in the ancestral vertebrate. *Journal of Molecular Evolution* 71(3):219-230.

Kiss, A.J., Mirarefi, A.Y., Ramakrishnan, S., Zukoski, C.F., Devries, A.L., Cheng, C.H. (2004). Cold-stable eye lens crystallins of the Antarctic nototheniid toothfish *Dissostichus mawsoni* Norman. *Journal of Experimental Biology* 207(26): 4633–4649.

Koop, B.F., von Schalburg, K.R., Leong, J., Walker, N., Lieph, R., Cooper, G.A., *et al.* (2008). A salmonid EST genomic study: genes, duplications, phylogeny and microarrays. *BMC Genomics* 9:545.

Laprairie, R.B., Denovan-Wright, E.M., Wright, J.M. (2016a). Divergent evolution of *cis*-acting peroxisome proliferator-activated receptor elements that differentially control the tandemly duplicated fatty acid-binding protein

genes, *fabp1b.1* and *fabp1b.2*, in zebrafish. *Genome* 59(6):403-412.

Laprairie, R.B., Denovan-Wright, E.M., Wright, J.M. (2016). Subfunctionalization of peroxisome proliferator response elements accounts for retention of duplicated *fabp1* genes in zebrafish. *BMC Evolutionary Biology* 16(1):147.

Lien, S., Koop, B.F., Sandve, S.R., Miller, J.R., Kent, M.P., Nome, T., *et al.* (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature* 533:200-205.

Litt, M., Kramer, P., LaMorticella, D.M., Murphey, W., Lovrien, E.W., Weleber, R.G. (1998). Autosomal dominant congenital cataract associated with a missense mutation in the human alpha crystallin gene CRYAA. *Human Molecular Genetics* 7:471–474.

Livak, K.J., Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25(4):402-408.

Mahler, B., Chen, Y., Ford, J., Thiel, C., Wistow, G., Wu, Z. (2013). Structure and dynamics of the fish eye lens protein, γM7-crystallin. *Biochemistry* 52(20):3579-3587

Mao, L., Bryantsev, A.L., Chechenova, M.B., Shelden, E.A. (2005). Cloning, characterization, and heat stress-induced redistribution of a protein homologous to human hsp27 in the zebrafish *Danio rerio*. *Experimental Cell*

Research 306(1):230-241.

Mao, L., Shelden, E.A. (2006). developmentally regulated gene expression of the small heat shock protein *Hsp27* in zebrafish embryos. *Gene Expression Patterns* 6(2):127-133.

McFall-Ngai, M.J., Horwitz, J. (1990). A comparative study of the thermal stability of the vertebrate eye lens: Antarctic ice fish to the desert iguana. *Experimental Eye Research* 50(6):703-709.

Meyer, A., Schartl, M. (1999). Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Current Opinion in Cell Biology* 11(6):699-704.

Meyer, A., Van de Peer, Y. (2005). From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays* 27(9):937-945.

Morimoto, R.I. (1998). Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. *Genes and Development* 12:3788-3796.

Narberhaus, F. (2002). α-Crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network. *Microbiology and Molecular Biology Reviews* 66(1):64-93.

Narum, S.R., Campbell, N.R. (2010). Sequence divergence of heat shock genes within and among 3 Oncorhynchids. *Journal of Heredity* 101(1):107-112.

Ohno, S. (1970). Evolution by gene duplication. Springer-Verlang, Heidelberg, Germany.

Ohno, S., Wolf, U., Atkin, N.B. (1968). Evolution from fish to mammals by gene duplication. *Hereditas* 59:169–187.

Pan, F.M., Chang, W.C., Chao, Y.K., Chiou, S.H. (1994). Characterization of γ-crystallins from a hybrid teleostean fish: multiplicity of isoforms as revealed by cDNA sequence analysis. *Biochemical and Biophysical Research Communications* 202(1):527-534.

Parmar, M.B., Wright, J.M. (2013). Comparative genomic organization and tissue-specific transcription of the duplicated *fabp7* and *fabp10* genes in teleost fishes. *Genome* 56(11):691-701.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in in real-time RT-PCR. *Nucleic Acids Research* 29(9):e45.

Posner, M., Hawke, M., LaCava, C., Prince, C. J., Bellanco, N. R., & Corbin, R. W. (2008). A proteome map of the zebrafish (*Danio rerio*) lens reveals similarities between zebrafish and mammalian crystallin expression. *Molecular Vision 14*:806–814.

Rise, M.L, Douglas, S.E., Sakhrani, D., Williams, J., Ewart, K.V., Rise, M., Davidson, W.S., Koop, B.F., Devlin, R.H. (2006). Multiple microarray platforms utilized for hepatic gene expression profiling of GH transgenic

coho salmon with and without ration restriction. *Journal of Molecular Endocrinology* 37(2):259-282.

Rise, M.L., Jones, S.R.M., Brown, G.D., von Schalburg, K.R., Davidson, W.S., Koop, B.F. (2004a). Microarray analyses identify molecular biomarkers of Atlantic salmon macrophage and hematopoietic kidney response to *Piscirickettsia salmonis* infection. *Physiological Genomics* 20:21-35.

Rise, M.L., von Schalburg, K., Brown, G.D., Devlin, R.H., Mawer, M.A., Kuipers, N., *et al.* (2004b). Development and application of a salmonid EST database and cDNA microarray: datamining and interspecific hybridization characteristics. *Genome Research* 14:478-490.

Rozen, S., Skaletsky, H.J. (2000). Primer3 on the WWW for general users and for biologist programmers. Bioinformatics Methods and Protocols: Methods in Molecular Biology (Eds. Krawetz, S., Misener, S.), Humana Press, New Jersey, 365-386.

Rupik, W., Jasik, K., Bembenek, J., Widlak, W. (2011). The expression patterns of heat shock genes and proteins and their role during vertebrate's development. *Comparative Biochemistry and Physiology, Part A: Molecular and Integrative Physiology* 159(4):349-366.

Sidow, A. (1996). Gen(om)e duplications in the evolution of early vertebrates. *Current Opinion in Genetics and Development* 6:715–722.

Slingsby, C., Wistow, G.J., Clark, A.R. (2013). Evolution of crystallins for role in the vertebrate eye lens. *Protein Science* 22(4):367-380.

Steinke, D., Hoegg, S., Brinkmann, H., Meyer, A., (2006). Three rounds (1R/2R/3R) of genome duplications and the evolution of the glycolytic pathway in vertebrates. *BMC Biology* 4:16-30.

Steinke, D., Salzburger, W., Meyer, A. (2006b). Novel relationships among ten fish model species revealed based on a phylogenomic analysis using ESTs. *Journal of Molecular Evolution* 62:722-784.

Taylor, J.S., Braasch, I., Frickey, T., Meyer, A., Van de Peer, Y. (2003). Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Research* 13:382-390.

Thirumaran, A., Wright, J.M. (2014). Fatty acid-binding protein (*fabp*) genes of spotted green pufferfish (*Tetradon nigroviridis*): comparative genomics and spatial transcriptomic regulation. *Genome* 57(5):289-301.

Tuttle, A.M., Gauley, J., Chan, N., Heikkila, J.J. (2007). Analysis of the expression and function of the small heat shock protein gene, *hsp27*, in *Xenopus laevis* embryos. *Comparative Biochemistry and Physiology, Part A: Molecular and Integrative Physiology* 147(1):112-121.

Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett, D.E. Jr., *et al.* (1997). Characterization of the yeast transcriptome. *Cell* 88(2):243-51.

von Schalburg, K.R., Rise, M.L., Cooper, G.A., Brown, G.D., Gibbs, A.R., Nelson, C.C., *et al.* (2005). Fish and chips: various methodologies demonstrate utility of a 16,006-gene salmonid microarray. *BMC Genomics* 6:126.

von Schaburg, K.R., Cooper, G.A., Leong, J., Robb, A., Lieph, R., Rise M.L., Davidson, W. S., Koop, B.F. (2008a). Expansion of the genomics research on Atlantic salmon *Salmo salar* L. project (GRASP) microarray tools. *Journal of Fish Biology* 72:2051-2070.

von Schalburg, K.R., Yazawa, R., de Boer, J., Lubieniecki, K.P., Goh, B., Straub, C.A., Beetz-Sargent, M.R., Robb, A., Davidson, W.S., Devlin, R.H., Koop, B.F. (2008b). Isolation, characterization and comparison of Atlantic and Chinook salmon growth hormone 1 and 2. *BMC Genomics* 9:522.

Wages, P., Horwitz, J., Ding, L., Corbin, R.W., Posner, M. (2013). Changes in zebrafish (*Danio rerio*) lens crystallin content during development. *Molecular Vision* 19: 408–417.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J. (2009). Jalview version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189-1191.

Wistow, G., Wyatt, K., David, L., Gao, C., Bateman, O., Bernstein, S., *et al.* (2005). γN-crystallin and the evolution of the βγ-crystallin superfamily in vertebrates. *Federation of European Biochemical Societies Journal* 272(9):2276-2291.

Wolfe, K.H. (2001). Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics* 2:333-341.

Xu, Q., Feng, C.Y., Hori, T.S., Plouffe, D.A., Buchanan, J.T., Rise, M.L. (2013). Family-specific differences in growth rate and hepatic gene expression in juvenile triploid growth hormone (GH) transgenic Atlantic salmon (*Salmo salar*). *Comparative Biochemistry and Physiology, Part D: Genomics and Proteomics* 8(4):317-33.

Xue, X., Feng, C.Y., Hixson, S.M., Johnstone, K., Anderson, D.M., Parrish, C.C., Rise, M.L. (2014). Characterization of the fatty acyl elongase (elovl) gene family, and hepatic elovl and delta-6 fatty acyl desaturase transcript expression and fatty acid responses to diets containing camelina oil in Atlantic cod (*Gadus morhua*). *Comparative Biochemistry and Physiology, Part B: Biochemistry and Molecular Biology* 175:9-22.

Zhao, H., Brown, P.H., Magone, M. T., Schuck, p. (2011). The molecular refractive function of lens γ-crystallins. *Journal of Molecular Biology* 411(3):680-699.

2.8 WEB REFERENCES

2.1 http://blast.ncbi.nlm.nih.gov/Blast.cgi/; BLAST: Basic Local Alignment Search Tool.

2.2 http://pbil.univ-lyon1.fr/cap3.php; CAP3 Sequence Assembly Program

2.3 http://web.uvic.ca/grasp/; Consortium for Genomics Research on All Salmon Project.

2.4 http://prosite.expasy.org/; ExPASy Prosite: Database of Protein Domains, Families, and Functional Sites.

2.5 http://web.expasy.org/translate/; ExPASy Translate Tool.

2.6 http://www.ncbi.nlm.nih.gov/nucest/; GenBank NCBI EST database.

2.7 http://www.ebi.ac.uk/Tools/msa/muscle/; MUSCLE Multiple Sequence Alignment.

2.8 http://frodo.wi.mit.edu/primer3/; Primer3 V.4.0.

2.9 http://www.uniprot.org/; UniProt Knowledgebase.

2.10 http://edis.ifas.ufl.edu/fa031; Water Quality: Ammonia in Aquatic Systems

2.11 http://ceqg-rcqe.ccme.ca/download/en/141; Water Quality: Canadian Water Quality Guidelines for the Protection of Aquatic Species.

2.12 https://www.dep.state.fl.us/labs/docs/unnh3disc.docx; Water Quality: Explanation of the Discrepancy between the Florida Department of

Environmental Protection's Un-ionized Ammonia Calculator (Version 2.1) and the EPA Text Table 3 Values in EPA 440/5-88-004

3. FUNCTIONAL GENOMICS STUDY OF ATLANTIC SALMON SAC FRY RESPONSES TO INCREMENTAL INCREASES IN WATER TEMPERATURE

3.1 ABSTRACT

The 32K cGRASP microarray was used to identify 11 genes that were more highly expressed in 85 dpf sac fry at 7 °C relative to 11 °C, 23 genes that were more highly expressed at 11 °C relative to 7 °C, 8 genes that were more highly expressed in 89 dpf sac fry at 7 °C relative to 15 °C, and 13 genes that were more highly expressed at 15 °C relative to 7 °C. QPCR analysis was used to examine the transcript expression of four microarray-identified, candidate hyperthermia-responsive biomarkers (*trypsin-1 precursor*, *chymotrypsin b*, *ferritin middle subunit*, and *ubiquitin*), as well as four *α-sHSP* paralogues during chronic incremental hyperthermia. *Trypsin-1 precursor*, *ferritin middle subunit*, and *ubiquitin*, as well as *α-sHSPs 1*, *3* and *4*, were shown by QPCR as responsive to the incremental hyperthermia conditions. The expression profiles of these genes provide information on the constitutive and hyperthermia-responsive expression of these transcripts during the late sac fry stage of development. The disruption of the constitutive expression of these transcripts could have implications for the fitness and adaptability of Atlantic salmon in relation to climate change.

3.2 INTRODUCTION

The sac fry stage of Atlantic salmon development is a crucial time in the life cycle of this species. Prior to hatch, the salmon embryo is separated from the environment by the extra-embryonic membranes (i.e., the enveloping layer and the chorion) which provide protection against both physical and chemical stressors (Finn 2007). After hatching, the sac fry remain in the substrate for a period, but then enter the water column to begin feeding (Scott and Scott 1988). It is at this stage that the salmon sac fry face multiple environmental stressors (e.g., immune challenges, changes in temperature and oxygen availability), as well as competition for resources and predation (Finn 2007). The environmental conditions encountered during the sac fry stage of development, and the ability of an individual to adapt to these changing environmental conditions, may play important roles in the survival and evolutionary fitness of an individual.

Hatch occurs in the spring, and the riffle sections of streams that are the favoured spawning grounds of Atlantic salmon can experience frequent thermal fluctuations as air temperature increases, and snow melt, as well as heavy rain, can infuse the system with copious amounts of fresh cool water. Developmental duration is temperature dependent, and one of the greatest environmental challenges facing the sac fry is changing water temperature. The 2007 Intergovernmental Panel on Climate Change's (IPCC) fourth assessment report predicted that the effects of climate change will influence fresh water more severely than marine systems

(Bogner *et al.* 2008). It is likely that Atlantic salmon sac fry may face increased exposure to temperature-related stress in the future. Thus, it has been predicted that Atlantic salmon, as well as other salmonids in the Northern hemisphere with similar life cycles and breeding strategies, will likely respond to climate change by: 1) a northern shift in population distribution; 2) spawning later in the summer (fall); and 3) accelerating embryonic and sac fry development (Jonsson and Jonsson 2009). In addition, expected changes in climate portend increased salmonid mortality rates and disease susceptibility (Jonsson and Jonsson 2009).

The stress response in fish has been studied under various paradigms, and in response to diverse stressors (Aluru and Vijayan 2009; Tort 2011). Of particular interest to the current study are previous experiments conducted on teleost fish to identify heat stress-responsive genes (which can function as molecular biomarkers), to elucidate the underlying mechanisms involved in heat stress response, and to examine the potential effects of heat stress on fitness. Previous studies have determined that teleost transcriptome responses to heat stress are related to the severity and duration of exposure (Logan and Somero 2011; Long *et al.* 2012). cDNA encoding genes functionally annotated as "protein folding" were identified as dysregulated in the liver, head kidney and skeletal muscle of Atlantic cod (*Gadus morhua*) exposed to heat stress, and an immune-associated transcript that may be involved in the mechanism of heat-stress-induced immunosuppression (TLR22) was down-regulated in head kidney (Hori *et al.* 2010). Atlantic salmon fry reared under

hyperthermic conditions (>12 °C after first-feeding) had accelerated growth rate, but this was concomitant with higher instances of spinal column deformities (Ytteborg *et al.* 2010). Heart rate and cardiovascular activity can also be affected by heat stress, as rainbow trout sac fry exposed to increased temperatures (5, 10, 12, and 15 °C) had elevated heart rates (Mirkovic and Rombough 1998). Nonetheless, despite these previous studies, little is known about the Atlantic salmon sac fry transcript expression responses to elevated temperature, and how these gene expression changes may influence their ability to adapt to elevated temperature.

Sequencing of the Atlantic salmon genome (Davidson *et al.* 2010; Lien *et al.* 2016) suggests divergence from the zebrafish (the closest related organism with a draft genome) approximately 200 million years ago (m.y.a.) (Jaillon *et al.* 2004). As discussed in Chapter 2 of this thesis, genome duplication events are thought to have occurred early in the ray-finned fish lineage (320-350 m.y.a.) and in the salmonid lineage (approximately 80 m.y.a.; Lien *et al.* 2016) (Allendorf and Thorgaard 1984; Koop *et al.* 2008; Ohno 1970). Extant salmonids are considered to be pseudotetraploid (Allendorf and Thorgaard 1984; Danzmann *et al.* 2006; Koop *et al*. 2008), and have many duplicated genes (paralogues) that may diverge and develop new functions, and thus, potentially play a role in evolutionary processes (i.e., most commonly by neofunctionalization) (Li *et al.* 2007; Ohno 1970; Ohno *et al.*1968; Ravi and Venkatesh 2008; Wolfe 2001). Paralogues may also sub-functionalize [dividing the functions of the ancestral genes among duplicated ‘copies’ (Force *et al.*

1999)], accumulate mutations and become silenced (i.e., become pseudogenes), or retain ancestral functions (Ohno 1970; Jaillon *et al.* 2004).

Over the past decade, salmonid genomics projects have generated resources for functional genomics research including genome sequences, cDNA libraries, expressed sequence tag (EST) databases, and DNA microarrays (e.g., Adzhubei *et al.* 2007; Hagen-Larsen *et al.* 2005; Koop *et al*. 2008; Lien *et al.* 2016; Rise *et al*. 2004b; von Schalburg *et al*. 2005). For example, there is a large knowledge base for bioinformatic and functional genomic studies involving Atlantic salmon, with 498,245 Atlantic salmon nucleotide ESTs archived in NCBI's Taxonomy Browser (accessed August 26th, 2016; Web Reference 3.2). In addition, a 44K cGRASP salmonid oligonucleotide array with an associated dataset on the developmental transcriptome has been developed (Jantzen *et al.* 2011). Microarray platforms are high-throughput analytical tools that enable better understanding of a species' transcript expression as affected by various biological processes such as embryogenesis or responses to stressors (e.g., Rise *et al.* 2004a; Jantzen *et al.* 2011; Hori *et al.* 2012; Booman *et al.* 2011). In this study, bioinformatics, molecular biology, and functional genomics tools and techniques were employed to study gene expression responses of Atlantic salmon sac fry exposed to chronic incremental hyperthermia, and to identify hyperthermia-responsive biomarkers that may be involved in coping with hyperthermia at this critical developmental stage. In addition, I report on the transcript expression profiles of four previously identified

Atlantic salmon *α-sHSP* paralogues (Chapter 2) in sac fry exposed to a chronic incremental increase in ambient water temperature. At the time that the microarray-based component of the current research was conducted, the 32K cGRASP salmonid cDNA microarray platform represented the most comprehensive (i.e., had the best coverage of the transcriptome) publicly available microarray for Atlantic salmon. Therefore, the 32K platform was used for the current global transcript expression analyses.

3.3 METHODS

3.3.1 Experimental design, water quality, and sampling procedure

Atlantic salmon sac fry that remained in the three flow-through tanks at the Marine Institute's (MI) aquaculture facility after sampling for the research presented in Chapter 2 were used in this chapter. Due to financial constraints, only one diurnal incubator was available for this research. Given this limitation, the best possible experimental design was devised to allow comparable oxygen and other water quality parameters in the control and experimental tanks. Sac fry were divided into two groups: the control group remained in the three flow-through tanks, and the experimental group was transferred to the incubator. As described in Chapter 2 of this thesis, the control condition was set up in three flow-through tanks with water from the MI well. Each tank was equipped with an air stone, and the water temperature remained at 7.4 °C for the duration of the experiment. The experimental

condition was set up in the incubator with sac fry divided among triplicate 2L glass beakers within the incubator (70 individuals in each of three beakers to be sampled for gene expression analysis). Additionally, a set of triplicate 2L glass beakers were set up as mortality tanks (100 individuals in each of three beakers to be used to determine cumulative mortality). Each beaker was equipped with an air stone and filled with water obtained from the MI well (with static half-volume water changes every 48 h, immediately after taking water quality measurements). The incubator was initially set to 7.4 °C (and sac fry in the incubator were acclimated at 7.4 °C for 48 h prior to initial sampling), and the temperature was increased by 1 °C every 24 h, in order to complete the thermal experiment during the sac fry developmental stage. Temperature in the experimental tank increased from 7.4-21.4 °C in order to encompass the 12 °C upper thermal tolerance range for sac fry (DeCola 1970; Elliott 1991; Peterson *et al.* 1977), and approach the 22.5 °C upper thermal tolerance range for the following life stage (i.e., fry at/after first feeding) (DeCola 1970; Elliott 1991). Mortalities were removed every 24 h.

Water temperature, un-ionized ammonia levels, and percent dissolved oxygen saturation (DO) were recorded every 24 h from 81-95 dpf, as described previously (see Chapter 2, section 2.3.2.). Un-ionized ammonia (NH_3 in $mg \bullet L^{-1}$) levels were calculated from the total ammonia nitrogen (TAN) [using a LaMotte SMART 2 Colorimeter (LaMotte Company, Chestertown, MD)] as described in Chapter 2, section 2.3.2, of this thesis using the following calculations (Emerson *et al.* 1975;

Web Reference 3.5, 3.6, and 3.7):

Equation 1: $NH_3 = (TAN)\ (f)$ (LaMotte constant)

The LaMotte constant is 1.216 [obtained by dividing the molar mass of ammonia (17.031 $g \bullet mol^{-1}$) into the molar mass of ammonia as nitrogen (14.007 $g \bullet mol^{-1}$)] (Web Reference 3.7) and *f* is the fraction of total ammonia that is un-ionized, as described in Equation 2 (Emerson *et al.* 1975; Web Reference 3.6 and 3.7).

Equation 2: $f = 1 / [10^{(pK-pH)} + 1]$

The pH was 7.5 for the duration of the experiment. In Equation 2, pK refers to the equilibrium constant (Emerson *et al.* 1975; Web Reference 3.6 and 3.7):

Equation 3: pK = 0.09018 + 2729.2 / (temperature °C + 273.2)

Water quality and cumulative mortality statistical analysis methods were outlined and conducted by Dr. Albert Caballero-Solares, as follows: changes in mortality, and temperature, oxygen, and ammonia levels in water over time and between conditions were statistically analyzed following Wei et al. (2012) recommendations. Briefly, experimental data were subjected to two-way analysis of the variance (ANOVA), including *Time* and *Condition* as fixed factors. *Time* factor comprised fifteen levels corresponding to all sampling time points (i.e., 81-95 dpf); *Condition* factor had two levels: *control* and *experimental*. If both factors showed a significant interaction, then the effects of both were evaluated jointly through one-way ANOVA. To this aim, a new experimental factor was created by combining the

different *Time* and *Condition* levels. Thus, this new factor (*Time* x *Condition*) presented thirty levels (e.g., 81 dpf/control, 81 dpf/experimental, 82 dpf/control, etc.). For pairwise comparisons between *Time* x *Condition* combinations, the one-way ANOVA was followed by Tukey's (homogeneity of variances among groups) or Games-Howell (variances not homogenous across groups) post-hoc tests. If no significant interaction was found between *Time* and *Condition*, then factor effects were analyzed separately through one-way ANOVA. For pairwise comparisons, significant changes over *Time* were further studied using Tukey's or Games-Howell post-hoc tests. In contrast, significant *Condition* effects were further investigated by Student t-test (data normally distributed) or Mann-Whitney U test (data not normally distributed). The significance level was established at $p < 0.05$.

Sampling occurred at 81 dpf, 85 dpf, 89 dpf and 93 dpf (Figure 2.1). Although the control and experimental fish were time-matched (dpf), exposure to incremental hyperthermia in the experimental condition would cause a discrepancy in degree days (dd) as the experiment proceeded (i.e., in the control condition relative to the experimental condition, respectively: 81 dpf = 599.4 dd and 599.4 dd; 85 dpf = 629 dd and 639 dd; 89 dpf = 658.6 dd and 694.6 dd; and 93 dpf = 673.4 dd and 766.2 dd). At these time points, three sac fry were sampled from each triplicate experimental and control tank, placed in individual nuclease-free microcentrifuge tubes, immediately flash-frozen in crushed dry ice, and stored in a -80 °C freezer. All utensils and tools were disinfected with iodine, rinsed with distilled water, and

treated with RNase Away (Molecular BioProducts, San Diego, CA) before use and between individuals. Sac fry were kept in the dark at all times, except during sampling.

3.3.2 RNA extraction and purification

Individual sac fry were ground to a fine powder using baked (220 °C for 5 h) ceramic mortars and pestles (see Chapter 2, section 2.3.4.), and RNA was extracted from the homogenized tissue using TRIzol reagent (Invitrogen, Carlsbad, CA) following the manufacturer's instructions with the previously described modifications to the procedure (see Chapter 2, section 2.3.4.). At the end of the TRIzol-based RNA isolation procedure, each total RNA pellet (representing RNA from an individual fish) was re-suspended in 50 µl of nuclease-free water (Invitrogen, Carlsbad, CA). The RNA samples were then stored at -80 °C until DNase 1 digestion and column purification.

Total RNA (30 µg from each individual) was digested with 6.8 Kunitz units of DNase 1 (RNase-Free DNase set, QIAGEN, Mississauga, ON; following the manufacturer's methods, as described in section 2.3.4) to degrade genomic DNA. Each RNA sample was column-purified (RNeasy MinElute Cleanup Kit, QIAGEN; following the manufacturer's instructions) using an RNeasy MinElute (QIAGEN) spin column to remove any traces of DNase 1, salt, and other impurities, and eluted with 20 µl of nuclease-free water (Invitrogen). Finally, the cleaned RNA samples

were stored at -80 °C until further use.

Individuals were selected for the microarray experiment (see section 3.3.3.) based on high RNA yield and quality. Total RNA quantity, quality, and integrity were assessed for pre-cleaned and column-purified samples using A260/A280 and A260/A230 NanoDrop ND1000 UV spectrophotometry (with an accepted absorbency ratio range of 1.8-2.2) and 1% agarose gel electrophoresis. Gels were used to electrophoretically separate 1 µg of RNA per individual, and were stained with ethidium bromide and run in 1X TAE buffer at 100 volts for 30 min alongside a 1 Kb Plus DNA ladder (Invitrogen). Gels were visualized with a UV transilluminator in a G:BOX (Syngene, Frederick, ML).

3.3.3 Microarray hybridization and data acquisition

Two direct comparison microarray experiments were designed using the cGRASP 32K (salmonid) cDNA microarray (Koop *et al*. 2008) to identify transcripts that were differentially expressed between control and experimental age-matched fish. Four 6 µg pools of total RNA [consisting of 1 µg from each of 6 individuals (two individuals per triplicate tank) with high RNA yield and purity] were used to generate labelled targets for the microarray experiment. The first microarray experiment, with three technical replicates including one dye-swap (to account for dye bias), compared global transcript expression of pooled late sac fry sampled at 11 °C (85 dpf) relative to age-matched pooled control late sac fry at 7 °C (85 dpf)

(Figure 3.4). The second experiment, with four technical replicates, including two dye-swaps, compared global transcript expression of pooled late sac fry sampled at 15 °C (89 dpf) relative to age-matched pooled control late sac fry at 7 °C (89 dpf) (Figure 3.4). These two time points were selected for direct comparison in order to identify potential heat-responsive transcripts. The Array 900 Detection Kit (Genisphere, Hatfield, PA) (Cy3 and Cy5) and SuperScript II (Invitrogen) were used for microarray target synthesis and subsequent hybridization following the manufacturer's instructions (as in Hall *et al.* 2011). Microarray hybridization (at 50 °C) was conducted using 1 μg of pooled total RNA for synthesis of each target following the manufacturers' instructions, with the previously described changes and additions to the procedure (see Chapter 2, section 2.3.5.). Arrays were scanned immediately at 10 μm resolution with laser power set to 90% to obtain fluorescent images (Cy3 emitting at 570 nm, fluorescing green; Cy5 emitting at 670 nm, fluorescing red) using ScanArray Express (PerkinElmer). PMT values were adjusted while scanning the arrays in order to obtain approximately equivalent signal intensities between the Cy3 and Cy5 channels for the three technical replicate microarrays (PMT 67-70 for Cy3; PMT 78-81 for Cy5) comparing the 85 dpf experimental and control fish, and for the four technical replicate microarrays (PMT 65-69 for Cy3; PMT 78-81 for Cy5) comparing the 89 dpf experimental and control fish.

3.3.4 Microarray data analysis

ImaGene v7.5 software (Biodiscovery, El Segundo, CA) was used to extract signal intensity data from TIFF images corresponding to each channel of each hybridized, scanned microarray, and the data were analyzed using GeneSpring 7.3 (Silicon Genetics) (see Chapter 2, section 2.3.6.). Four gene lists were obtained from the microarray studies, representing genes that were greater than 2-fold differentially expressed [higher or lower at 7 °C compared with 11 °C (85 dpf), and higher and lower at 7 °C compared with 15 °C (89 dpf)] on all slides of a study. Associated gene names and gene ontology (GO) terms (biological process, molecular function, and/or cellular component) and associated numeric identifiers were obtained (see Chapter 2, section 2.3.6.) from the NCBI BLAST server and the UniProt Knowledgebase (Kb) (Web Reference 3.1 and 3.4) to provide functional annotation of genes that were differentially expressed between hyperthermic and control sac fry. Genes of interest (GOI) were selected (section 3.3.5.), in addition to the four sHSP paralogues discussed previously in this thesis (Chapter 2), for further analysis by QPCR.

3.3.5 QPCR primer design and quality testing

One GOI was selected from each of the four microarray-generated lists of genes that were differentially expressed during the thermal experiment. Informative features that were represented multiple times in a list, or multiple times among lists, and that were annotated with defence-relevant or digestive functions, were selected

as GOI. Defence-relevant features could illustrate some of the adaptive strategies employed by Atlantic salmon sac fry upon exposure to elevated temperatures, while features associated with digestive functions were of interest because the sac fry were undergoing yolk reabsorption, and the effect of elevated temperature on digestive transcripts could influence fitness as the sac fry approached first-feeding. Primer3 (Web Reference 3.3; Rozen and Skaletsky 2000) was used to design primers for each GOI using the application's default settings, with the exception that primer size was set to 20 bases. Complementary DNA (cDNA) targets for QPCR primer quality testing were synthesized from 1 µg of pooled DNase 1-treated RNA, with three high-quality (i.e., high quantity, purity, and integrity, as determined by gel electrophoresis and NanoDrop spectrophotometry) individuals from the experimental condition from each time point (81, 85, 89, and 93 dpf) (n = 9 for each of the four time points) contributing equally (see Chapter 2, section 2.3.9.). Candidate primer sets were quality tested by running a five-point 1:3 serial dilution (starting with 10x diluted cDNA) standard curve with technical triplicate reactions in 96-well format (see Chapter 2, section 2.3.9.). All QPCR was conducted using the 7500 Fast Real-Time PCR System (Applied Biosystems 7500 Fast 2.0). Fluorescence and baseline thresholds were set automatically. Multiple primer sets were also designed and quality tested for six candidate normalizers identified as stably transcribed [background corrected, Lowess normalized (BCLN) expression ratio between 0.8 and 1.2 fold on all slides of the microarray study including the dye-swaps]. In

addition, primer sets for four candidate normalizers that, while not identified by the microarray results, are commonly used in the Rise Lab, were also quality tested (data not shown). Candidate GOI and normalizer primer sets were quality tested following the methods in Rise *et al.* 2008, and as described previously in Chapter 2, section 2.3.9. The paralogue-specific primers that were designed for each of the four sHSP paralogues in Chapter 2 of this thesis (section 2.3.9) were also used in the current study to determine the expression of these transcripts as affected by the incremental temperature change. Primer sets that passed QPCR primer quality testing, and that were used for QPCR analysis, are shown in Table 3.1.

3.3.6 cDNA synthesis and QPCR assays of selected genes of interest

To synthesize cDNA, 1 µg of high-quality (as determined by gel electrophoresis and NanoDrop spectrophotometry) DNase 1-treated and column-purified RNA from each individual whole larva was reverse transcribed using MMLV-RT (Invitrogen) as per the manufacturer's instructions, with a final reaction volume of 20 µl (see Chapter 2, section 2.3.9.; Rise *et al.* 2008). A “no reverse transcription” (“no-RT”) control was also run. This was a cDNA synthesis reaction from a pool of RNA from all individuals (1 µg RNA per individual, pooled prior to cDNA synthesis), that was set up as described in Chapter 2, section 2.3.9., with the modification that the reaction omitted RTase. The no-RT controls were run in triplicate for each primer target set, and no amplification was observed. This

Table 3.1 Paralogue-specific primer sets used for QPCR analysis of four paralogous *α-sHSP* transcripts[1] and four Atlantic salmon GOI identified by microarray analysis.

Gene of Interest	Primer Name	Primer Length (bases)	Sequence (5' – 3')	Amplicon Size (bases)	% Efficiency [2]	Up-regulated In:
***S.salar* α-sHSP 1**	2sHSP1_L	20	AAACAGGCCAAAGCAGAAGA	115	95.1	N/A
	2sHSP1_R	20	GAGCAGGCAGGTCAATTTTC			
***S.salar* α-sHSP 2**	2sHSP2_L	20	AAACAGGCCAAAGCAGAAGA	141	101.9	N/A
	2sHSP2_R	20	ACAGGTCAAGGCTGTCAGGT			
***S.salar* α-sHSP 3**	8sHSP3_L	20	CCACCTGCTTGACAGACAGA	164	87.9	N/A
	8sHSP3_R	20	TATGCAACGACTCCACCAAA			
***S.salar* α-sHSP 4**	1sHSP4_L	20	ATATGCAACGACTGCACCAA	148	107.9	N/A
	1sHSP4_R	20	AGCTCCTCTGGGGAGAAGTC			
CB492538 (trypsin-1 precursor)	6trypsin1_L	20	ATCGTCGGAGGGTATGAGTG	235	82.9	7 °C compared to 11 °C
	6trypsin1_R	20	AGGAGCTGTAGTTGGGGTGA			
CB491721 (ferritin middle subunit)	6ferritin-ms1_L	20	CAGGTGGAGGCCATTAAGAA	201	85.5	11 °C compared to 7 °C
	6ferritin-ms1_R	20	GCCAAAAGCCAGCAGAATAG			
CB515463 (chymotrypsin b)	3chymotryp4_L	20	GGACAGCTTGATCAGGGAGA	125	103.9	7 °C compared to 15 °C
	3chymotryp4_R	20	GAGCATAAGAAGGGCAGTGG			
EG770691 (ubiquitin)	1ubiquitin1_L	20	GGTATAATGCCCAGGGAGGT	101	96.7	15 °C compared to 7 °C
	1ubiquitin1_R	20	TCGTTCCTATGAGACCAGCA			
40S ribosomal protein S3 (normalizer[3])	40S_S3a_L	20	GATTGTCGACCCTTTCTCCA	138	81.2	N/A
	40S_S3a_R	20	ACACGTCCCTTCAGACCATC			

[1]Primer design for the four *α-sHSP* paralogues is described in section 2.3.9. of this thesis.

[2]%Efficiency values were used to calculate relative quantity (RQ).

[3]Although many candidate normalizers were tested, and others passed QPCR primer quality testing (data not shown), *40S ribosomal protein S3* had the most stable transcription (all C_t values within a range of < 0.5 cycles) across all time points and for all individuals in the study.

confirmed the absence of genomic DNA in the samples.

While the microarray experiment utilized RNA pools, QPCR was conducted using individual RNA samples to assess biological variability of transcript expression of selected genes. The four microarray-identified GOI and four previously described *α-sHSP* paralogues were normalized to the expression of the stably transcribed (threshold cycle (C_t) range of < 0.5 cycles across all individual samples in the study) normalizer *40S S3 ribosomal protein S3*, which had been developed and previously used as a normalizer by members of the Rise Lab. Each individual cDNA sample (biological replicate) was run in technical triplicate for the target GOI and the normalizer on the same 96-well plate, and amplification of the transcripts was conducted as described previously in Chapter 2, section 2.3.9. A control amplification reaction (a cDNA template synthesized by reverse transcription from a 1 µg pool of DNase 1-treated, column-purified RNA, consisting of equal contributions from three high-quality individual samples from the 89 dpf experimental condition) was run on each plate of the study, in order to link plates and to evaluate technical variability between plates. Linker sample C_t values were within a 0.5 cycle range on all plates in the QPCR experiment. C_t values were used to calculate the relative quantity (RQ) of each transcript using the $2^{-\Delta\Delta Ct}$ quantification method (Livak and Schmittgen 2001; Pfaffl 2001) with the calculated amplification efficiencies of the primer sets (Table 3.1). Data were corrected for technical replicate outliers using an accepted C_t range of 0.5 cycles for technical triplicate reactions. RQ

values are expressed as means ± SEM. The individual with the lowest normalized expression for a particular GOI (regardless of the condition or sampling time point of that individual) was used as the calibrator sample (i.e., RQ = 1.00). Fold-change was determined relative to the initial sampling point (control sac fry, 7 °C, 81 dpf). Statistical analyses (SigmaPlot 12 for Windows) consisted of: 1) two-way ANOVA, with group (control or experimental) and age (dpf) as main effects; and 2) the Student-Newman-Keuls post-hoc test to identify significantly different sample means ($p < 0.05$) (Pérez-Casanova *et al.* 2008).

3.4 RESULTS

3.4.1 Experimental paradigm and water quality monitoring at the MI

The control and experimental conditions were run in a flow-through system and an incubator, respectively, as previously explained. The water quality and mortality in tanks for both of these conditions were carefully observed.

The two-way ANOVA found DO was significantly different between the control and experimental conditions, but was not significantly different over time within either condition (Table A.3). No significant interactions were observed between *Condition* and *Time* (Table A.3). The nonparametric comparison of the two *Condition* groups (independently of *Time*) confirmed that differences in DO were significant between the control and experimental tanks (Table A.4). As shown in

Figure 3.1, oxygen saturation levels in the experimental condition was lower than in the control condition.

A significant interaction was found between *Time* and *Condition* for the NH_3 levels; NH_3 was significantly different between the control and experimental conditions, and also increased significantly over time in the experimental tanks while remaining consistent (no significant differences over time) in the control tank (Tables A.5 and A.6; Figure 3.1). Significant differences between between the control and experimental conditions were observed as early as the 83 dpf time point (Figure 3.1).

Both DO and NH_3 increase in the experimental tanks showed a fluctuating pattern resulting from static half-volume water changes every 48 h (as described in the methods, section 3.3.1). Cumulative mortality was significantly different between the control and experimental treatments (*Condition*) (Tables A.8 and A.9) as of the 87 dpf time point (Figure 3.2). Cumulative mortality increased significantly over the experimental period (*Time*) (Tables A.8 and A.9) and mortality accumulated at a higher rate in the experimental tanks. This increase in mortality in the experimental tanks, and the significant differences between the control and experimental treatments, could be a result of the thermal regime, but could also be confounded by the higher concentration of un-ionized NH_3 in the experimental tanks.

It is important to consider the different set-up conditions between the control and experimental tanks, as well as the difference in water quality, when interpreting the results of this study; in particular, un-ionized NH_3 may have been a confounding

Figure 3.1 Water quality during Atlantic salmon late sac fry development (81-95 dpf) under chronic incremental hyperthermic and control (held at 7 °C) conditions. Water originated from the MI well. Water was tested every 24 h for temperature (°C), dissolved oxygen (% saturation), and un-ionized ammonia (NH_3) concentration ($mg \bullet L^{-1}$). The sections highlighted in grey (7 °C vs. 11 °C at 85 dpf and 7 °C vs. 15°C at 89 dpf) indicate time points examined by the microarray experiment. A significant interaction was found between *Time* and *Condition* for the NH_3 levels (see Table A.5 and A.6). Therefore, post-hoc tests were conducted on all *Time* x *Condition* combinations, rather than separately for each factor (see section 3.3.1). Different letter case assignment is intended to facilitate the interpretation of the figure: upper case letters correspond to experimental condition data points, while lower case letters correspond to control condition data points. Different letters indicate significant ($p < 0.05$) differences between NH_3 data points..

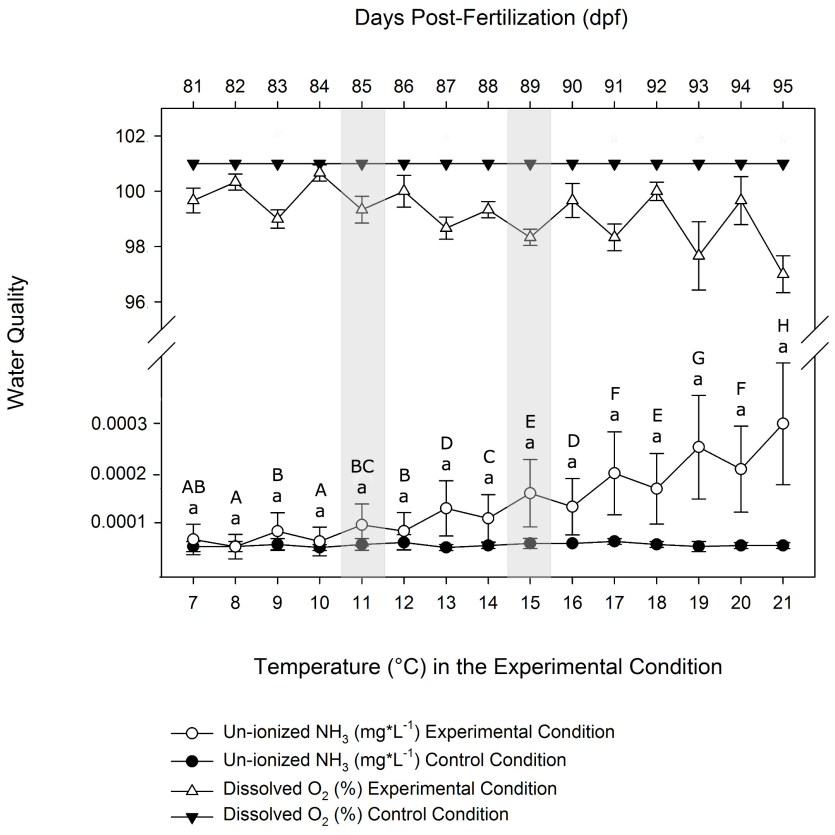
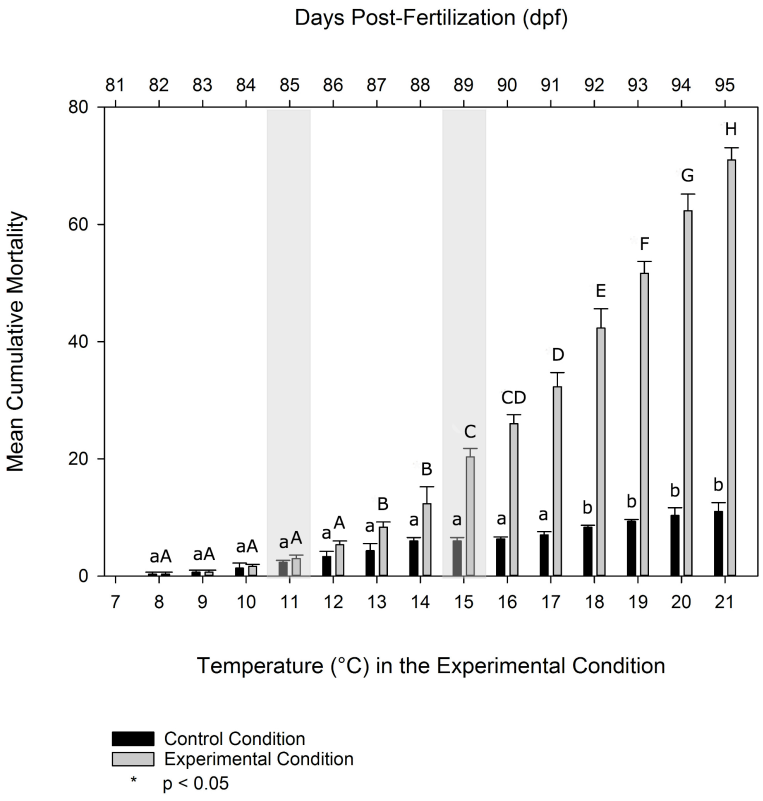

Days Post-Fertilization (dpf)
81 82 83 84 85 86 87 88 89 90 91 92 93 94 95
Water Quality
102
100
98
96
0.0003
0.0002
0.0001
AB a
A a
B a
A a
BC a
B a
D a
C a
E a
D a
F a
E a
G a
F a
H a
7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Temperature (°C) in the Experimental Condition
Un-ionized NH3 (mg*L-1) Experimental Condition
Un-ionized NH3 (mg*L-1) Control Condition
Dissolved O2 (%) Experimental Condition
Dissolved O2 (%) Control Condition

Figure 3.2 Cumulative mortality of sac fry under control conditions (held at 7 °C) and when exposed to chronic incremental hyperthermia (ambient temperature increased by 1 °C every 24 hours, from 7-21 °C). The control tanks were flow-through, whereas the experimental tanks were beakers with air stones in an incubator. Cumulative mortality data were obtained from the triplicate control tanks and from the triplicate mortality tanks maintained in the incubator for the duration of the experiments. The sections highlighted in grey (7 °C vs. 11 °C at 85 dpf and 7 °C vs. 15°C at 89 dpf) indicate time points examined by the microarray experiment. A significant interaction was found between *Time* and *Condition* for cumulative mortality (see Table A.4). Therefore, post-hoc tests were conducted on all *Time* x *Condition* combinations, rather than separately for each factor (see section 3.3.1). Different letter case assignment is intended to facilitate the interpretation of the figure: upper case letters correspond to experimental condition data points, while lower case letters correspond to control condition data points. Different letters indicate significant ($p < 0.05$) differences between cumulative mortality data points.


Days Post-Fertilization (dpf)
81 82 83 84 85 86 87 88 89 90 91 92 93 94 95
Mean Cumulative Mortality
0 20 40 60 80
aA aA aA aA aA
a A a B a B a C a CD a D a E b F b G b H
7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Temperature (°C) in the Experimental Condition
Control Condition
Experimental Condition
* p < 0.05

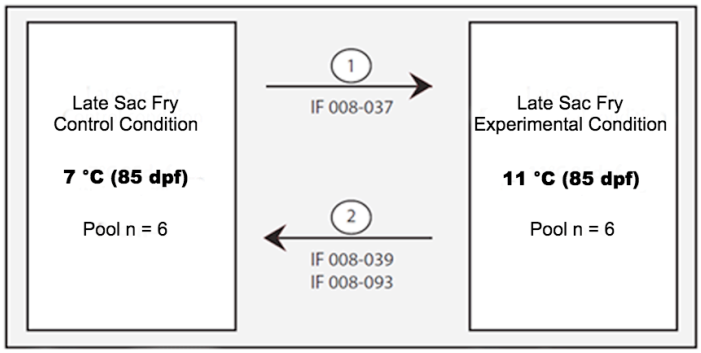
factor affecting transcript expression of the experimental sac fry. The limitations of this experiment, including the differences in water quality between the control and experimental conditions, are discussed in Section 3.5.1.

3.4.2 Functional annotation and identification of GOI for QPCR analysis

Lists of informative genes were obtained from the microarray study using the 32K cGRASP salmonid array platform to identify Atlantic salmon sac fry transcripts that respond to chronic incremental hyperthermia (Figure 3.3). In the first comparison, between age-matched sac fry at 7 °C and 11 °C (85 dpf), 11 microarray features were more than 2-fold down-regulated in 11 °C sac fry compared to 7 °C sac fry (Table 3.2), while 23 features were identified as more than 2-fold up-regulated in 11 °C compared to 7 °C sac fry (Table 3.3). In the second comparison, between age-matched sac fry at 7 °C and 15 °C (89 dpf), 8 features were identified as more than 2-fold down-regulated in 15 °C compared to 7 °C sac fry (Table 3.4), and 13 features were more than 2-fold up-regulated in 15 °C compared to 7 °C sac fry (Table 3.5). The best (i.e. lowest E-value) BLASTx hit with an associated gene or protein name (i.e., not “hypothetical” or “predicted”), as well as functional annotation from the UniProt Kb (Web references #3.5), was obtained for each differentially expressed cDNA feature on the arrays (Tables 3.2-3.5). There was considerable “internal validation” (i.e., multiple same-named features) within the microarray-identified

Figure 3.3 Global gene expression compared (A) between late sac fry at 11 °C (85 dpf) and age-matched control late sac fry at 7 °C (85 dpf), and (B) between late sac fry at 15 °C (89 dpf) and age-matched control late sac fry at 7 °C (89 dpf) using the cGRASP 32K salmonid microarray. Six individuals from each of the time points were equally represented in 6 µg RNA pools used for target synthesis. Targets were fluorescently labeled with either Cy3 or Cy5. The batch and slide number of technical replicates used in this study are indicated, as well as the number of genes that passed threshold and were reproducibly greater than 2-fold up-regulated or down-regulated (see Tables 3.1-3.4) in each comparison (A or B).

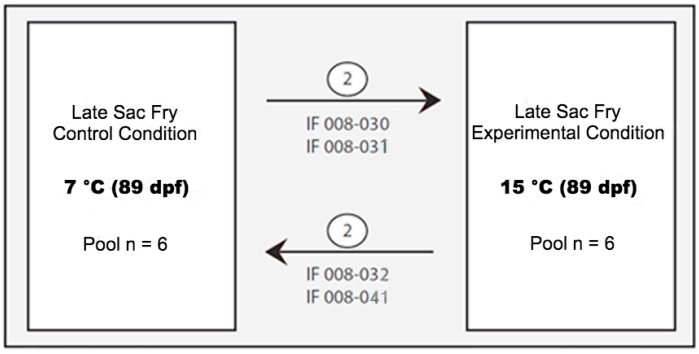
A



The expression of 11 genes was > 2-fold down-regulated at 11 °C (85 dpf) compared 7 °C (85 dpf) (Table 3.2)

The expression of 23 genes was > 2-fold up-regulated at 11 °C (85 dpf) compared 7 °C (85 dpf) (Table 3.3)

B



The expression of 8 genes was > 2-fold down-regulated at 15 °C (89 dpf) compared 7 °C (89 dpf) (Table 3.4)

The expression of 13 genes was > 2-fold up-regulated at 15 °C (89 dpf) compared 7 °C (89 dpf) (Table 3.5)

Table 3.2 Eleven reproducibly informative[1] genes down-regulated in Atlantic salmon late sac fry at 11 °C (85 dpf) compared to age-matched control late sac fry at 7 °C (85 dpf).

EST Accession Number	Gene Name of Best[2] BLAST[3] Hit	Length (%)[4]	E-value	Functional Annotation[5] of Best BLAST Hit	GO Identifier[6]	Mean Fold Change	SEM
CB492958	Chymotrypsinogen 2-like protein (AAT45254.1; *Sparus aurata*)	184 (83)	3e-78	BP: proteolysis; MF: peptidase activity, catalytic activity, hydrolase activity, serine-type peptidase activity, serine-type endopeptidase activity.[7]	GO:0006508 GO:0008233 GO:0003824 GO:0016787 GO:0008236 GO:0004252	4.97	2.38
CA050427	Unknown	N/A	N/A	N/F	N/F	2.97	0.74
CA050822	Chymotrypsin B (NP_001134565.1; *Salmo salar*)	178 (100)	4e-104	BP: proteolysis; MF: serine-type peptidase activity, serine-type endopeptidase activity, hydrolase activity, peptidase activity.	GO:0006508 GO:0008236 GO:0004252 GO:0016787 GO:0008233	2.95	0.51
DY696138	Chromosome 7 genomic contig, GRCh37 reference primary assembly (NT_007933.15; *Homo sapiens*)[3]	61 (100)	4e -24	N/F	N/F	2.93	0.63
CA054856	Unknown	N/A	N/A	N/F	N/F	2.74	0.72
CK990598	Elastase-1 (Q7SIG3.1; *Salmo salar*)	122 (82)	4e-46	BP: proteolysis; MF: metal ion binding, serine-type peptidase activity, serine-type endopeptidase activity, peptidase activity, hydrolase activity, catalytic activity; CC: extracellular region.	GO:0006508 GO:0046872 GO:0008236 GO:0004252 GO:0008233 GO:0016787 GO:0003824 GO:0005576	2.57	0.48
CA059987	Deoxyribonuclease gamma precursor (BT057320.1; *Salmo salar*)	45 (84)	3e-15	BP: DNA catabolic process; MF: deoxyribonuclease activity.	GO:0006308 GO:0004536	2.50	0.12
CB512116	Chymotrypsin B (NP_001134565.1; *Salmo salar*)	178 (100)	3e-101	BP: proteolysis; MF: serine-type peptidase activity, serine-type endopeptidase activity, hydrolase activity, peptidase activity.	GO:0006508 GO:0008236 GO:0004252 GO:0016787 GO:0008233	2.40	0.25
CA051324	Unknown	N/A	N/A	N/F	N/F	2.26	0.35
CB492538	**Trypsin-1 precursor (ACI67104.1; *Salmo salar*)**	**146 (97)**	**2e-82**	**BP: proteolysis; MF: serine-type peptidase activity, serine-type endopeptidase activity, hydrolase activity, peptidase activity.**	**GO:0006508 GO:0008236 GO:0004252 GO:0016787 GO:0008233**	**2.24**	**0.29**

CB516178	Trypsin IA precursor [*Salmo salar*] (NP_001117183.1; *Salmo salar*)	190 (100)	3e-137	BP: proteolysis, digestion; MF: catalytic activity, serine-type peptidase activity, serine-type endopeptidase activity, hydrolase activity, peptidase activity, metal ion binding.	GO:0006508 GO:0007586 GO:0003824 GO:0008236 GO:0004252 GO:0016787 GO:0008233 GO:0046872	2.14	0.17

[1]2-fold or greater down-regulation in 11 °C compared to age-matched 7 °C Atlantic salmon late sac fry on all three slides of the study (including one dye-swap).

[2]The best BLASTx or BLASTn hit (putative orthologue), defined as having the lowest E-value (≤ 1e-10) and an associated gene or protein name (e.g., excluding "hypothetical" or "predicted"), is shown. Species: *Homo sapiens* (human), *Salmo salar* (Atlantic salmon), and *Sparus aurata* (gilt-head sea bream). BLAST reports were collected on May 27th, 2011, and reflect the entries collected in the GenBank non-redundant (nr) protein and nucleic acid sequence databases to this date. N/A: not applicable. GOI that were further analyzed by QPCR are indicated in bold font.

[3]Indicates that a BLASTn hit was used where no significant BLASTx hit was retrieved.

[4]Length of aligned region, in amino acid residues, and the percent identity, indicating the number of exactly matched residues between the query sequence (i.e., the microarray-identified feature) and its best BLAST hit.

[5]Functional annotation associated with *Salmo salar*. N/F: no functional annotation found for best BLAST hit or any putative orthologues. Gene ontology (GO) categories: biological process (BP), molecular function (MF) and cellular component (CC).

[6]GO identifiers are listed in the same relative order as the functional annotations (GO terms) with which they are associated.

[7]Functional annotations associated with other teleost species.

Table 3.3 Twenty-three reproducibly informative[1] genes up-regulated in Atlantic salmon late sac fry at 11 °C (85 dpf) compared to age-matched control late sac fry at 7 °C (85 dpf).

EST Accession Number	Gene Name of Best[2] BLAST[3] Hit	Length (%)[4]	E-value	Functional Annotation[5] of Best BLAST Hit	GO Identifier[6]	Mean Fold Change	SEM
CA059307	Response gene to complement 32 protein (rgc32), mRNA (NM_001146584.1; *Salmo salar*)	537 (99)	0	N/F	N/F	7.27	4.03
EG849106	Claudin 28b (ACK86563.1; *Oncorhynchus mykiss*)	175 (99)	5e-68	MF: structural molecule activity; CC: tight junction, plasma membrane, integral to membrane, cell junction, membrane.[7]	GO:0005198 GO:0005923 GO:0005886 GO:0016021 GO:0030054 GO:0016020	3.98	0.88
CB508046	Complement C1q-like protein 4 (NP_001140082.1; *Salmo salar*)[3]	79 (48)	1e -10	N/F	N/F	3.68	0.48
CK990704	Zgc:162944 protein (AAI35079.1; *Danio rerio*)	78 (57)	1e-17	N/F	N/F	3.58	0.69
EG814386	Transcriptome shotgun assembly (TSA): 7119 mRNA sequence (EZ770453.1; *Oncorhynchus mykiss*)[3]	572 (94)	0	N/F	N/F	3.20	0.85
EG782284	Unknown	N/A	N/A	N/F	N/F	3.14	0.39
EG813888	Unknown	N/A	N/A	N/F	N/F	3.03	0.80
EG768537	TSA: 98 mRNA sequence (EZ763432.1; *Oncorhynchus mykiss*)[3]	422 (89)	1e -148	N/F	N/F	2.86	0.52

CK990553	Amino-peptidase N (AF012465.1; *Pseudopleuro-nectes americanus*)	86 (80)	1e-33	BP: proteolysis; MF: aminopeptidase activity, metallopeptidase activity, zinc ion binding, hydrolase activity.	GO:0006508 GO:0004177 GO:0008237 GO:0008270 GO:0016787	2.74	0.54
CA040158	Insulin-like growth factor-binding protein 6 precursor putative mRNA, complete cds, clone ssal-rgf-514-311 (BT045207.1; *Salmo salar*) [3]	722 (100)	0	BP: regulation of cell growth; MF: insulin-like growth factor binding; CC: extracellular region.[8]	GO:0001558 GO:0005520 GO:0005576	2.74	0.04
CK991031	Ferritin, middle subunit (ACI68839.1; *Salmo salar*)	90 (87)	2e-37	BP: cellular iron ion homeostasis; MF: ferric iron binding.	GO:0006879 GO:0008199	2.56	0.29
CA064175	Physical map contig 483, genomic sequence (EU481821.1; *Salmo salar*)[3]	434 (85)	1e-116	N/F	N/F	2.55	0.43
EG888926	Unknown	N/A	N/A	N/F	N/F	2.50	0.12
CA051643	Zgc:162944 protein (AAI35079; *Danio rerio*)	92 (64)	3e-27	N/F	N/F	2.46	0.16
CB507396	Ferritin, middle subunit (ACI68639.1; *Salmo salar*)	126 (100)	3e-70	BP: cellular iron ion homeostasis; MF: ferric iron binding, metal ion binding.	GO:0006879 GO:0008199 GO:0046872	2.44	0.21
CK990310	Ferritin, middle subunit (ACN09837.1; *Salmo salar*)	97 (96)	8e-49	BP: oxidation reduction, cellular iron ion homeostasis, iron ion transport; MF: transition metal ion binding, ferric iron binding, binding, oxidoreductase activity, metal ion binding.	GO:0055114 GO:0006879 GO:0006826 GO:0046914 GO:0008199 GO:0005488 GO:0016491 GO:0046872	2.43	0.31
CB498370	Ferritin, middle subunit (ACO08179.1; *Oncorhynchus mykiss*)	113 (100)	8e-60	BP: oxidation reduction, cellular iron ion homeostasis, iron ion transport; MF: transition metal ion binding, ferric iron binding, binding, oxidoreductase activity, metal ion binding.[7]	GO:0055114 GO:0006879 GO:0006826 GO:0046914 GO:0008199 GO:0005488 GO:0016491 GO:0046872	2.40	0.23

EG864166	Collagen type X alpha 1 (ADG29155.1; *Epinephelus coioides*)	153 (80)	3e-66	N/F	N/F	2.37	0.30
CA058458	Zinc finger CCCH domain-containing protein 15 (NP_001139846.1; *Salmo salar*)	152 (88)	3e-58	MF: nucleic acid binding, zinc ion binding, metal ion binding.	GO:0003676 GO:0008270 GO:0046872	2.31	0.41
EG854579	Zinc finger protein 391 (NP_001140120.1; *Salmo salar*)	54 (61)	3e-07	MF: nucleic acid binding, zinc ion binding; CC: intracellular.	GO:0003676 GO:0008270 GO:0005622	2.30	0.36
EG872650	Junction plakoglobin putative mRNA, complete cds, clone ssal-rgf-538-157 (BT045980.1; *Salmo salar*) [3]	369 (98)	0	MF: binding.[8]	GO:0005488	2.26	0.15
CA056697	TSA: 28238 mRNA sequence (EZ791572.1; *Oncorhynchus mykiss*)[3]	742 (88)	0	N/F	N/F	2.20	0.05
CB491721	**Ferritin, middle subunit (ACI69640.1; *Salmo salar*)**	**38 (97)**	**3e-15**	**BP: cellular iron ion homeostasis; MF: ferric iron binding.**	**GO:0006879 GO:0008199**	**2.18**	**0.19**

[1]2-fold or greater up-regulation in 11 °C compared to age-matched 7 °C Atlantic salmon late sac fry on all three slides of the study (including one dye-swap).

[2]The best BLASTx or BLASTn hit (putative orthologue), defined as having the lowest E-value (≤ 1e-10) and an associated gene or protein name (e.g., excluding "hypothetical" or "predicted"), is shown. Species: *Danio rerio* (zebrafish), *Epinephelus coioides* (orange-spotted grouper), *Homo sapiens* (human), *Oncorhynchus mykiss* (rainbow trout), *Pseudopleuronectes americanus* (winter flounder), and *Salmo salar* (Atlantic salmon). BLAST reports were collected on May 27th, 2011, and reflect the entries collected in the GenBank non-redundant (nr) protein and nucleic acid sequence databases to this date. N/A: not applicable. GOI that were further analyzed by QPCR are indicated in bold font.

[3]Indicates that a BLASTn hit was used where no significant BLASTx hit was retrieved.

[4]Length of aligned region, in amino acid residues, and the percent identity, indicating the number of exactly matched residues between the query sequence (i.e., the microarray-identified feature) and its best BLAST hit.

[5]Functional annotations associated with *Salmo salar*. N/F: no functional annotation found for best BLAST hit or any putative orthologues. Gene ontology (GO) categories: biological process (BP), molecular function (MF) and cellular component (CC).

[6]GO identifiers are listed in the same relative order as the functional annotations (GO terms) with which they are associated.

[7]Functional annotations associated with *Oncorhynchus mykiss* or other teleost species[(8)].

Table 3.4 Eight reproducibly informative[1] genes down-regulated in Atlantic salmon late sac fry at 15 °C (89 dpf) compared to age-matched control late sac fry at 7 °C (89 dpf).

EST Accession Number	Gene Name of Best[2] BLAST[3] Hit	Length (%)[4]	E-value	Functional Annotation[5] of Best BLAST Hit	GO Identifier[6]	Mean Fold Change	SEM
CA356917	rCG64175 (CH474107.1; *Rattus norvegicus*)	151 (47)	3e-27	BP: antigen processing and presentation, immune response, cell adhesion; CC: membrane, MHC class I protein complex, extracellular space.[7]	GO:0019882 GO:0006955 GO:0007155 GO:0016020 GO:0042612 GO:0005615	5.94	3.12
CB511019	Formin-binding protein 1 gene, partial sequence and TCR gamma locus region, clone 242N16 (EU221177.1; *Salmo salar*)[3]	254 (94)	2e -105	N/F	N/F	2.93	0.63
CB515463	**Chymotrypsin B (NP_001134565.1; *Salmo salar*)**	**192 (100)**	**3e-110**	**BP: proteolysis; MF: serine-type peptidase activity, serine-type endopeptidase activity, hydrolase activity, peptidase activity.**	**GO:0006508 GO:0008236 GO:0004252 GO:0016787 GO:0008233**	**2.91**	**0.30**
CB493886	HLA-B associated transcript 1 (bat1), mRNA (NM_001173819.1; *Salmo salar*)[3]	57 (94)	2e -14	BP: transmembrane transport, transport, amino acid transport; MF: amino acid transmembrane transporter activity; CC: integral to membrane, membrane.	GO:0055085 GO:0006810 GO:0006865 GO:0015171 GO:0016021 GO:0016020	2.86	0.34
CB516178	Trypsin IB (CAA49677.1; *Salmo salar*)	190 (100)	6e-111	BP: proteolysis, digestion; MF: peptidase activity, catalytic activity, serine-type endopeptidase activity, serine-type peptidase activity, hydrolase activity, metal ion binding; CC: extracellular region, extracellular space.	GO:0006508 GO:0007586 GO:0008233 GO:0003824 GO:0004252 GO:0008236 GO:0016787 GO:0046872	2.81	0.33
CB492538	Trypsin-1 precursor (ACI67104.1; *Salmo salar*)	146 (97)	2e-82	BP: proteolysis; MF: serine-type peptidase activity, serine-type endopeptidase activity, hydrolase activity, peptidase activity.	GO:0006508 GO:0008236 GO:0004252 GO:0016787 GO:0008233	2.62	0.31

EG818497	Phosphatidyleth-anolamine-binding protein 1 (NP_001135084.1; *Salmo salar*)	187 (100)	1e-109	BP: response to oxidative stress, response to heat, response to organic nitrogen, positive regulation of acetylcholine metabolic process, negative regulation of MAPKKK cascade, response to wounding, negative regulation of protein amino acid phosphorylation, response to organic substance, regulation of the force of heart contraction, response to stress, response to toxin, positive regulation of cAMP-mediated signalling, brain development, eating behaviour, positive regulation of mitosis, response to ethanol, MAPKKK cascade, spermatid development, response to corticosterone stimulus, response to organic cyclic substance, ageing, response to activity, response to calcium ion, response to drug, response to electrical stimulus, response to cAMP, regulation of neurotransmitter levels; MF: ATP binding, protein binding, mitogen-activated protein kinase binding, phosphatidylethanolamine binding, serine-type endopeptidase inhibitor activity, protein kinase binding, kinase binding, receptor binding, peptidase inhibitor activity, lipid binding, nucleotide binding; CC: cytoplasm, soluble fraction, synaptosome, axon terminus, extracellular space, Golgi apparatus, synaptic vesicle, neuronal cell body, mitochondrion, apical part of cell, cell surface, rough ER, mitochondrial outer membrane.[7]	GO:0006979 GO:0009408 GO:0010243 GO:0060409 GO:0043409 GO:0009611 GO:0001933 GO:0010033 GO:0002026 GO:0006950 GO:0009636 GO:0043950 GO:0007420 GO:0042755 GO:0045840 GO:0045471 GO:0000165 GO:0007286 GO:0051412 GO:0014070 GO:0007568 GO:0014823 GO:0051592 GO:0042493 GO:0051602 GO:0051591 GO:0001505 GO:0005524 GO:0005515 GO:0051019 GO:0008429 GO:0004867 GO:0019901 GO:0019900 GO:0005102 GO:0030414 GO:0008289 GO:0000166 GO:0005737 GO:0005625 GO:0019717 GO:0043679 GO:0005615 GO:0005794 GO:0008021 GO:0043025 GO:0005739 GO:0045177 GO:0009986 GO:0005791 GO:0005741	2.41	0.25

CB494595	CIT protein (AAI23443.1; *Bos taurus*)	101 (45)	7e-10	MF: small GTPase regulator activity, metal ion binding.[8]	GO:0005083 GO:0046872	2.32	0.09

[1]2-fold or greater down-regulation in 15 °C compared to age-matched 7 °C Atlantic salmon late sac fry on all four slides of the study (including two dye-swaps).

[2]The best BLASTx or BLASTn hit (putative orthologue), defined as having the lowest E-value (≤ 1e-10) and an associated protein or gene name (e.g., excluding "hypothetical" or "predicted"), is shown. Species: *Bos taurus* (cattle), *Salmo salar* (Atlantic salmon), and *Rattus norvegicus* (rat). BLAST reports were collected on May 27th, 2011, and reflect the entries collected in the GenBank non-redundant (nr) protein and nucleic acid sequence databases to this date. N/A: not applicable. GOI that were further analyzed by QPCR are indicated in bold font.

[3]Indicates that a BLASTn hit was used where no significant BLASTx hit was retrieved.

[4]Length of aligned region, in amino acid residues, and the percent identity, indicating the number of exactly matched residues between the query sequence (i.e., the microarray-identified feature) and its best BLAST hit.

[5]Functional annotations associated with *Salmo salar*. N/F: no functional annotation found for best BLAST hit or any putative orthologues. Gene ontology (GO) categories: biological process (BP), molecular function (MF) and cellular component (CC).

[6]GO identifiers are listed in the same relative order as the functional annotations (GO terms) with which they are associated.

[7]Functional annotations associated with *Rattus norvegicus*, or *Homo sapiens*[(8)].

Table 3.5 Thirteen reproducibly informative[1] genes up-regulated in Atlantic salmon late sac fry at 15 °C (89 dpf) compared to age-matched control late sac fry at 7 °C (89 dpf).

EST Accession Number	Gene Name of Best[2] BLAST[3] Hit	Length (%)[4]	E-value	Functional Annotation[5] of Best BLAST Hit	GO Identifier[6]	Mean Fold Change	SEM
EG770691	**Ubiquitin (ACI69104.1; *Salmo salar*)**	**93 (100)**	**2e-46**	**N/F**	**N/F**	**25.10**	**16.48**
CA042089	Ependymin precursor (ACM09141.1; *Salmo salar*)	67 (100)	5e-33	BP: cell-matrix adhesion; MF: calcium ion binding; CC: extracellular region.	GO:0007160 GO:0005509 GO:0005576	23.85	20.54
CA039104	Progastricsin mRNA, complete cds (AF275939; *Salvelinus fontinalis*)[3]	557 (91)	0	BP: proteolysis; MF: peptidase activity, aspartic-type endopeptidase activity, hydrolase activity.[7]	GO:0006508 GO:0008233 GO:0004190 GO:0016787	23.63	20.68
CB505565	Type-2 ice-structuring protein (NP_001134281.1; *Salmo salar*)	158 (100)	3e-80	MF: binding, sugar binding.	GO:0005488 GO:0005529	7.98	1.92
CB509781	Ubiquitin-conjugating enzyme E2Q family member 2 (UBE2Q2), transcript variant 2, mRNA (Q8WVN8; *Homo sapiens*)[3]	80 (86)	2e -13	BP: regulation of protein metabolic process, post-translational protein modification; MF: ATP binding, ligase activity, small conjugating protein ligase activity, ubiquitin-protein ligase activity, nucleotide binding; CC: cytoplasm.[8]	GO:0051246 GO:0043687 GO:0005524 GO:0016874 GO:0019787 GO:0004842 GO:0000166 GO:0005737	4.79	2.20
EG789481	Tripartite motif-containing protein 16 (ACI34046.1; *Salmo salar*)	73 (49)	8e-12	MF: zinc ion binding, metal ion binding, protein binding; CC: intracellular.	GO:0008270 GO:0046872 GO:0005515 GO:0005622	4.41	1.19
CB503375	Type-2 ice-structuring protein (NP_001134281.1; *Salmo salar*)	158 (94)	2e-75	MF: binding, sugar binding.	GO:0005488 GO:0005529	4.07	1.01
EG869543	Interferon-induced guanylate-binding protein 1 (ACN10585.1; *Salmo salar*)	212 (59)	3e-44	MF: GTP binding, GTPase activity.	GO:0005525 GO:0003924	3.56	1.15

CA038770	C-type MBL-2 protein (NP_001117900.1; *Oncorhynchus mykiss*)	43 (95)	6e-19	MF: binding, sugar binding.[9]	GO:0005488 GO:0005529	3.36	0.24
EG860689	Transcobalamin 1 (NP_001122207.1; *Danio rerio*)	111 (42)	6e-19	BP: cobalamin transport; MF: cobalamin binding.[10]	GO:0015889 GO:0031419	3.25	0.42
EG881574	Alpha-N-acetyl-galactosaminidase (ACN10976.1; *Salmo salar*)	416 (100)	0	BP: metabolic process, carbohydrate metabolic process; MF: hydrolase activity, hydrolyzing O-glycosyl compounds, catalytic activity, hydrolase activity, acting on glycosyl bonds, cation binding, hydrolase activity.	GO:0008152 GO:0005975 GO:0004553 GO:0003824 GO:0016798 GO:0043169 GO:0016787	3.24	0.43
CB501157	BAC CH214-363E24 Atlantic Salmon BAC Library complete sequence (AC203446.12; *Salmo salar*)[3]	183 (88)	3e -52	N/F	N/F	2.78	0.53
CK990704	Zgc:162944 protein (AAI35079.1; *Danio rerio*)	78 (57)	1e-17	N/F	N/F	2.62	0.34

[1]2-fold or greater up-regulation in 15 °C compared to age-matched 7 °C Atlantic salmon late sac fry on all four slides of the study (including two dye-swaps).

[2]The best BLASTx or BLASTn hit (putative orthologue), defined as having the lowest E-value (≤ 1e-10) and an associated protein or gene name (e.g., excluding "hypothetical" or "predicted"), is shown. Species: *Danio rerio* (zebrafish), *Homo sapiens* (human), *Salmo salar* (Atlantic salmon), *Salvelinus fontinalis* (brook trout), and *Oncorhynchus mykiss* (rainbow trout). BLAST reports were collected on May 27th, 2011, and reflect the entries collected in the GenBank non-redundant (nr) protein and nucleic acid sequence databases to this date. N/A: not applicable. GOI that were further analyzed by QPCR are indicated in bold font.

[3]Indicates that a BLASTn hit was used where no significant BLASTx hit was retrieved.

[4]Length of aligned region, in amino acid residues, and the percent identity, indicating the number of exactly matched residues between the query sequence (i.e., the microarray-identified feature) and its best BLAST hit.

[5]Functional annotations associated with *Salmo salar*. N/F: no functional annotation found for best BLAST hit or any putative orthologues. Gene ontology (GO) categories: biological process (BP), molecular function (MF) and cellular component (CC).

[6]GO identifiers are listed in the same relative order as the functional annotations (GO terms) with which they are associated.

[7]Functional annotations associated with *Salvelinus fontinalis, Homo sapiens*[(8)], *Oncorhynchus mykiss* [(9)], or *Danio rerio*[(10)].

gene lists. For example, of the 11 informative features in Table 3.2 (down-regulated in 11 °C as compared to 7 °C in 85 dpf sac fry), *deoxyribonuclease gamma precursor* and *chymotrypsin b* were each represented by two different microarray features. In Table 3.3 (transcripts up-regulated in 11 °C as compared to 7 °C in 85 dpf sac fry), five different informative microarray features were identified as *ferritin middle subunit*. In Table 3.5 (transcripts up-regulated in 15 °C compared to 7 °C 89 dpf sac fry), type-2 ice-structuring protein is represented by two different microarray features.

3.4.3 QPCR analysis of transcript expression of four microarray-identified Atlantic salmon GOI and four Atlantic salmon *α-sHSP* paralogues

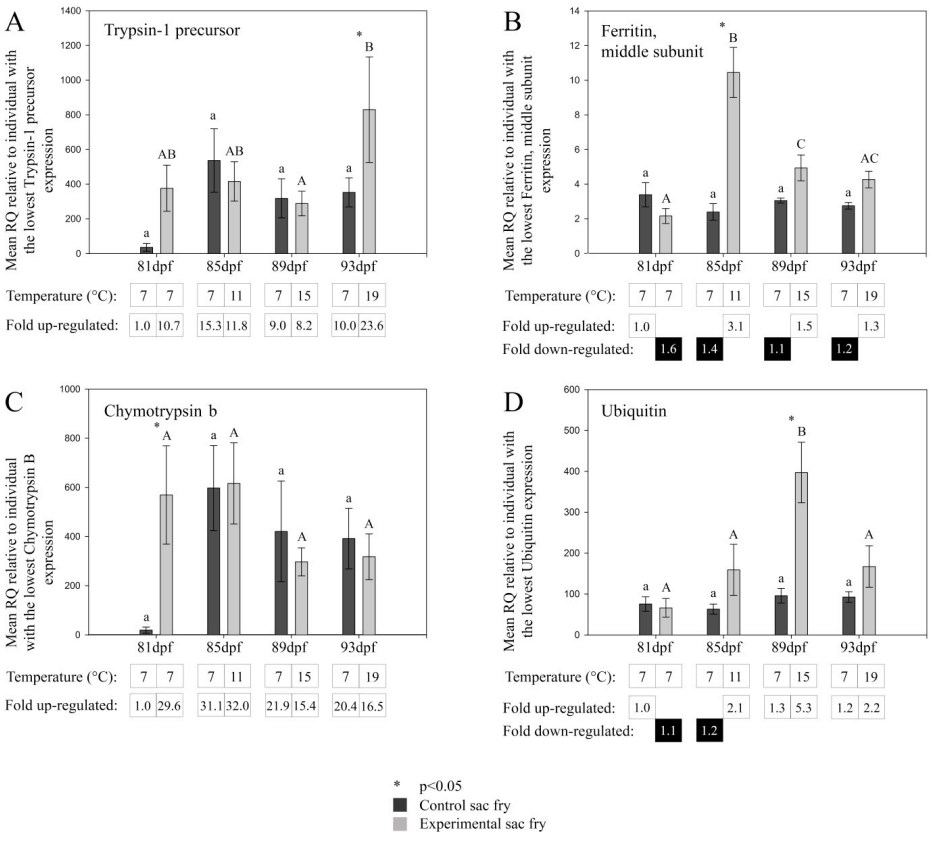
Four GOI (*trypsin-1 precursor*, EST accession number CB492538, Table 3.2 and Table 3.4; *chymotrypsin b*, CB515463, Table 3.4; *ferritin middle subunit*, CB 491721, Table 3.3; and *ubiquitin*, EG770691, Table 3.5) were selected from the microarry lists of informative genes for QPCR analysis. Down-regulated genes *trypsin-1 precursor* and *chymotrypsin b* were selected as GOI for QPCR based on their internal validation among lists of down-regulated genes, as well as their association with digestive functions. As the sac fry mature and complete yolk re-absorption, the down-regulation of digestive transcripts in response to the thermal regime could affect the success of the fry at first feeding. Up-regulated genes *ferritin*

middle subunit and *ubiquitin* were both selected as GOI for QPCR due to their functional relevance as stress-responsive genes. Additionally, *ferritin middle subunit* was highly represented within the list of genes up-regulated at 11 °C relative to 7 °C age-matched counterparts while *ubiquitin* had a mean fold change of 25.10 in 15 °C relative to 7 °C age-matched counterparts. Both *ferritin middle subunit* and *ubiquitin*, identified by the microarray as up-regulated in response to incremental increase in water temperature (Tables 3.4 and 3.6), were validated by the QPCR study (i.e., statistical analysis of RQ data indicated significantly higher expression (4.4 and 4.1-fold, respectively) in the experimental compared to the control condition at 85 (11 °C compared to 7 °C) and 89 dpf (15 °C compared to 7 °C), respectively) (Figure 3.4 B, D, Table A.11).

There were no significant differences observed in *trypsin-1 precursor* basal transcript expression in the control condition (7 °C) across the four sampling time points (Figure 3.4 A). A significant difference was observed in the expression of the *trypsin-1 precursor* transcript in sac fry in the experimental condition at 93 dpf compared to 89 dpf (2.9-fold up-regulated). A significant difference in *trypsin-1 precursor* transcript expression was also observed between the experimental condition and the age-matched control at 93 dpf (2.4-fold up-regulated) (Figure 3.4 A, Table A.11).

No significant differences, and very little biological variability, were observed in the constituitive transcript expression of *ferritin middle subunit* within

Figure 3.4 QPCR assessment of constitutive expression of four Atlantic salmon genes of interest (GOI) identified in the microarray experiment in whole late sac fry (81-93 dpf) (see Figure 3.3). Sac fry in the control group were reared at a constant average temperature of 7 °C, while those in the experimental group were reared in an incubator and exposed to chronic incremental hyperthermia (ambient temperature increased by 1 °C every 24 hours, from 7-21 °C). Values are mean ± SEM (n = 9). Relative quantity (RQ) was normalized to *40S ribosomal protein S3* and RQ values were calibrated to the individual with the lowest expression for each target. Groups without a letter in common are significantly ($p < 0.05$) different in either the control (lower case letters) or experimental (upper case letters) condition. Time points with a significant ($p < 0.05$) difference in transcript expression between age-matched control and experimental fish are indicated with an "*". Fold change was determined relative to the initial 7 °C control sampling time point at 81 dpf. Fold change in a white box with black text indicates up-regulation of expression at that time point relative to 81 dpf (control, 7 °C), while fold change in a black box with white text indicates down-regulation of expression at that time point relative to 81 dpf (control, 7 °C). All sac fry in the experiment still possessed yolk sacs, and had not yet commenced exogenous feeding.


A
Trypsin-1 precursor
Mean RQ relative to individual with the lowest Trypsin-1 precursor expression
81dpf
85dpf
89dpf
93dpf
Temperature (°C): 7 7 7 11 7 15 7 19
Fold up-regulated: 1.0 10.7 15.3 11.8 9.0 8.2 10.0 23.6
B
Ferritin, middle subunit
Mean RQ relative to individual with the lowest Ferritin, middle subunit expression
Temperature (°C): 7 7 7 11 7 15 7 19
Fold up-regulated: 1.0 3.1 1.5 1.3
Fold down-regulated: 1.6 1.4 1.1 1.2
C
Chymotrypsin b
Mean RQ relative to individual with the lowest Chymotrypsin B expression
Temperature (°C): 7 7 7 11 7 15 7 19
Fold up-regulated: 1.0 29.6 31.1 32.0 21.9 15.4 20.4 16.5
D
Ubiquitin
Mean RQ relative to individual with the lowest Ubiquitin expression
Temperature (°C): 7 7 7 11 7 15 7 19
Fold up-regulated: 1.0 2.1 1.3 5.3 1.2 2.2
Fold down-regulated: 1.1 1.2
* p<0.05
Control sac fry
Experimental sac fry

the control condition (7 °C) (Figure 3.4 B). *Ferritin middle subunit* was significantly up-regulated within the experimental condition at 85 dpf (11 °C) compared to 81, 89 and 93 dpf, and at 89 dpf (15 °C) compared to 81 and 85 dpf. The transcript expression of *ferritin middle subunit* was significantly up-regulated at 85 dpf (by 4.4-fold) in the experimental condition compared to the age-matched control counterpart (i.e., 11 °C compared to 7 °C) (Figure 3.4 B, Table A.11).

No significant difference in expression was observed for the *chymotrypsin b* transcript within either the control or the experimental group (Figure 3.4 C). This result was likely due to the high biological variability (indicated by large error bars) in the control condition at 85, 89, and 93 dpf (7 °C), and in the experimental condition at 81 (7 °C), 85 (11 °C), and 93 dpf (19 °C) (Figure 3.4 C). Increasing the number of biological replicates could have mitigated this issue. Nonetheless, *chymotrypsin b* transcript expression was significantly up-regulated (29.6-fold) between the experimental and the control condition at 81 dpf (both conditions at 7 °C) (Table A.11). It is possible that this dysregulation may have resulted from the transport of the experimental fish to their beakers, even though there was a 48 h acclimation period prior to the first sampling.

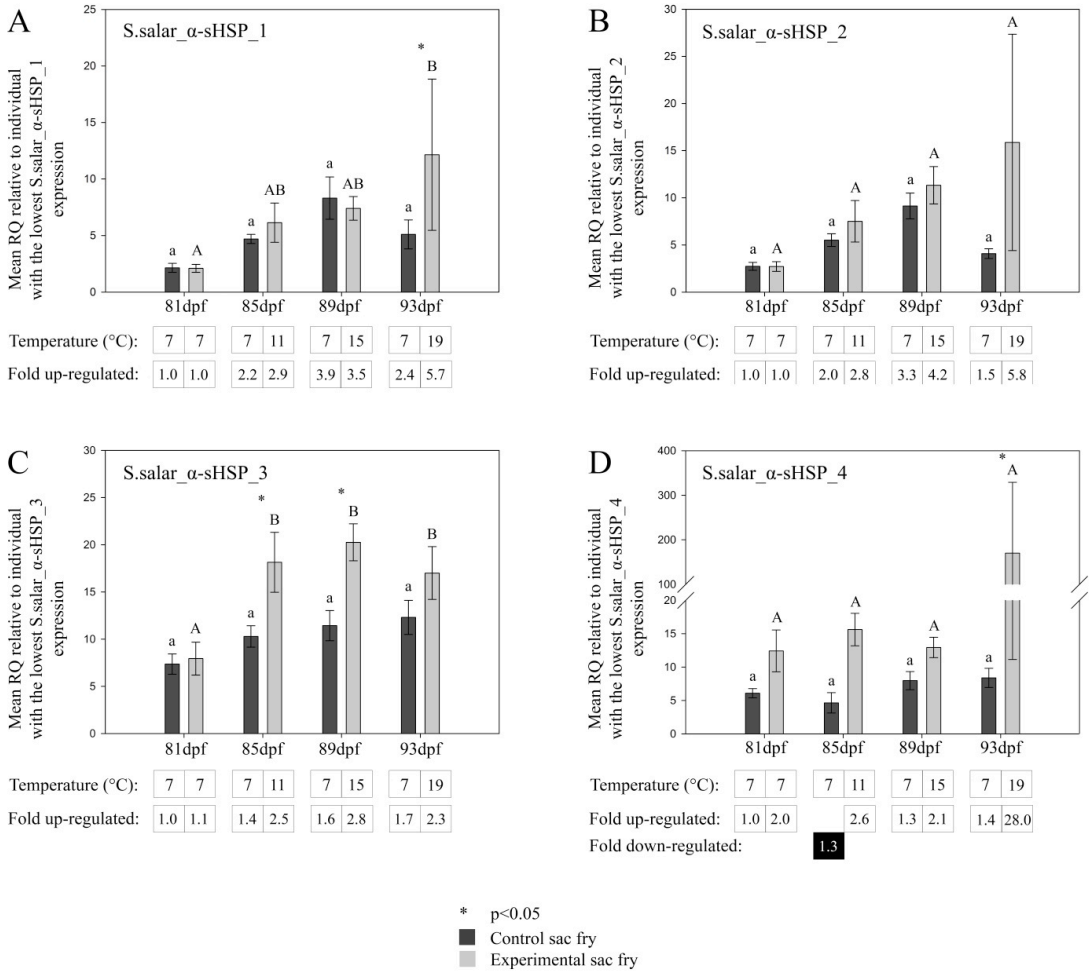
Constitutive transcript expression of *ubiquitin* remained consistant across all time points within the control condition (Figure 3.4 D). In the experimental condition, transcript expression of *ubiquitin* was significantly up-regulated at 89 dpf (15 °C) compared to all other time points (Figure 3.4 D). In addition, the *ubiquitin*

transcript was significantly up-regulated (by 4.1-fold) in the experimental condition at 89 dpf relative to age-matched control sac fry (i.e., 15 °C compared to 7 °C) (Figure 3.4 D, Table A.11).

The constitutive transcript expression profiles of sHSP 1 and sHSP 2, transcripts (which were 81% identical at the hypothetical amino acid level, Figure 2.6 B), appeared very similar (Figure 3.5 A, B) and no significant differences in expression were observed in the control group for these transcripts. Similarly, the pattern of expression was the same from 81 to 89 dpf in the experimental group. Although both sHSP 1 and sHSP 2 transcript expression appeared to steadily increase with increasing temperature in the experimental sac fry, this change in expression was only significant in sHSP 1 (5.7-fold up-regulated) at 93 (19 °C) compared to 81 dpf (7 °C) (Figure 3.5 A, Table A.11). Additionally, there was a significant up-regulation (by 2.4-fold) of sHSP 1 transcript expression in sac fry at 7 compared to 19 °C at 93 dpf. No significant differences in expression were observed for sHSP 2, possibly due to high biological variability, particularly in the 93 dpf (19 °C) experimental group (Figure 3.5 B).

The constitutive transcript expression profiles of sHSP 3 and sHSP 4 (83% identical at the hypothetical amino acid level; Figure 2.6 B), also appeared very similar in the control group, although, again, no significant differences in expression were observed (Figure 3.5 C, D). Transcript expression of sHSP 3 was significantly up-regulated at 85 (11 °C), 89 (15 °C) and 93 dpf (19 °C) relative to 81 dpf (7 °C)

Figure 3.5 QPCR assessment of constitutive and incremental hyperthermia-responsive expression of four Atlantic salmon *α-sHSP* paralogues in whole late sac fry (81 dpf-93 dpf). Sac fry in the control group were reared at a constant average temperature of 7 °C, while those in the experimental group were reared in an incubator and exposed to chronic incremental hyperthermia (ambient temperature increased by 1 °C every 24 hours, from 7-21 °C). Values are mean ± SEM (n = 9). Relative quantity (RQ) was normalized to *40S ribosomal protein S3* and RQ values were calibrated to the individual with the lowest expression for each target. Groups without a letter in common are significantly ($p < 0.05$) different in either the control (lower case letters) or experimental (upper case letters) condition. Time points with a significant ($p < 0.05$) difference in transcript expression between age-matched control and experimental fish are indicated with an "*". Fold change was determined relative to the initial 7 °C control sampling time point at 81 dpf. Fold change in a white box with black text indicates up-regulation of expression at that time point relative to 81 dpf (control, 7 °C), while fold change in a black box with white text indicates down-regulation of expression at that time point relative to 81 dpf (control, 7 °C). All sac fry in the experiment still possessed yolk sacs, and had not yet commenced exogenous feeding.


A
S.salar_α-sHSP_1
Mean RQ relative to individual with the lowest S.salar_α-sHSP_1 expression
81dpf 85dpf 89dpf 93dpf
Temperature (°C): 7 7 | 7 11 | 7 15 | 7 19
Fold up-regulated: 1.0 1.0 | 2.2 2.9 | 3.9 3.5 | 2.4 5.7
B
S.salar_α-sHSP_2
Mean RQ relative to individual with the lowest S.salar_α-sHSP_2 expression
81dpf 85dpf 89dpf 93dpf
Temperature (°C): 7 7 | 7 11 | 7 15 | 7 19
Fold up-regulated: 1.0 1.0 | 2.0 2.8 | 3.3 4.2 | 1.5 5.8
C
S.salar_α-sHSP_3
Mean RQ relative to individual with the lowest S.salar_α-sHSP_3 expression
81dpf 85dpf 89dpf 93dpf
Temperature (°C): 7 7 | 7 11 | 7 15 | 7 19
Fold up-regulated: 1.0 1.1 | 1.4 2.5 | 1.6 2.8 | 1.7 2.3
D
S.salar_α-sHSP_4
Mean RQ relative to individual with the lowest S.salar_α-sHSP_4 expression
81dpf 85dpf 89dpf 93dpf
Temperature (°C): 7 7 | 7 11 | 7 15 | 7 19
Fold up-regulated: 1.0 2.0 | 2.6 | 1.3 2.1 | 1.4 28.0
Fold down-regulated: 1.3
* p<0.05
Control sac fry
Experimental sac fry

(by 2.5, 2.8 and 2.3-fold, respectively) (Figure 3.5 C). The transcript expression of sHSP 3 was also significantly up-regulated in the experimental condition at 85 (11 °C) and 89 dpf (15 °C) compared to the age-matched control sac fry at 85 and 89 dpf (7 °C) (both by 1.8-fold) (Figure 3.5 C, Table A.11). sHSP 4 transcript expression in the experimental sac fry was not significantly affected by heat exposure (Figure 3.5 D). However, expression of sHSP 4 was significantly up-regulated in the experimental condition at 93 (19 °C) compared to time-matched control sac fry (7 °C) (Figure 3.5 D, Table A.11).

3.5 DISCUSSION

3.5.1 Limitations of the experiment

The experiment conducted in this chapter of my thesis utilized a sub-optimal experimental design, restricted by resource limitations and technical difficulties. However, several strategies have been utilized in an attempt to mitigate these limiting factors, and allow for the interpretation of the results.

The main concern affecting the interpretation of the results in this chapter is the difference between the control and experimental design. Specifically, the control condition occurred in flow-through tanks, while the experimental condition occurred in a diurnal incubator, with static water changes, and as a result, there were significant differences in the water quality between the experimental and control

conditions. Oxygen saturation was above 95% in both conditions, and were comparable from a physiological perspective (even though they were statistically significantly different between conditions). The level of un-ionized ammonia in the experimental tanks compared to the control tanks continued to climb over the course of the experiment. Additionally, mean cumulative mortality was significantly higher in the experimental condition relative to the control condition from 87 dpf onwards. In order to address these concerns, the following steps were taken: 1) this study did not analyse gene expression at the 21 °C time point, at which point mortality and un-ionized NH_3 were highest; 2) the microarray design examined the 11 °C and 15 °C experimental fish compared to their age-matched control counterparts. Although there was still a statistically significant difference in water quality between conditions at these time points, these differences were not biologically significant: un-ionized NH_3 and DO were within the recommended "no-harm" range (< 0.002 $mg \bullet L^{-1}$ and > 80%, respectively; Bergheim and Fivelstad 2014) for Atlantic salmon at the microarray time points (see Figures 3.1 and 3.2) as well as for the duration of the experiment . On the other hand, the microarray time points were thermal stress relevant, as the maximum thermal tolerance of Atlantic salmon sac fry is 12 °C (DeCola 1970; Elliott 1991; Peterson *et al.* 1977), and is likely the cause of the significant difference in mortality between the control and experimental conditions beginning at the 7 °C vs. 13 °C time point.

In addition to the disparity between the tank design and the water quality

between the experimental and control conditions, another concern was the conservative number of individuals sampled at each time point and the number of individuals selected for microarray analysis. Biological replication was not included in the microarray experimental designs; they included only technical replicate microarrays comparing RNA pools. Individual RNA samples were selected for inclusion in the pools based on concentration. However, both biological and technical replicates were incorporated into the QPCR experiment. The low number of individuals utilized in the RNA pools may have contributed to the number of genes identified by the microarray as being > 2-fold dysregulated and the failure to validate the microarray results for trypsin-1 precursor and chymotrypsin b using QPCR. A confounding factor was the poor quality of hybridization and persistent spatial effects on the microarrays, which caused the elimination one of the slides in the 7 °C vs. 11 °C dye-swap. Representing a greater percentage of the sampled fish by pooling more individuals for the microarray component of this study could have increased the likelihood of identifying genes that were validated by QPCR.

Although attempts have been made to justify the data acquired in this experiment, these limitations make it difficult to determine, with certainty, if temperature or water quality was the greater or sole influence on gene expression. The low number of biological replicates and the high variability in gene expression between individuals decreased statistical power, and prevented the identification of additional instances of statistical significance within the QPCR results. In light of

these issues, interpretation of the data, and the conclusions that follow, must be examined with consideration for the limitations described above.

3.5.2 Microarray-identified gene expression studies during chronic incremental hyperthermia in Atlantic salmon sac fry

The current study aimed to identify Atlantic salmon sac fry transcripts that respond to chronic incremental hyperthermia, with the understanding that these biomarkers (and associated QPCR assays) may be useful in future research on the impact of elevated temperature on wild and cultured early life stage salmon. Transcript expression of four microarray-identified GOI (*trypsin-1 precursor*, *chymotrypsin b*, *ferritin middle subunit*, and *ubiquitin*) was examined across four time points in the control (held at 7 °C) and experimental (7 °C, 11 °C, 15 °C, and 19 °C) condition. The remainder of this section of the discussion will focus on genes (and associated biological processes) that were identified as developmentally regulated and/or potentially hyperthermia-responsive in this study.

Although the *trypsin-1 precursor* and *chymotrypsin b* transcripts identified as differentially expressed by the microarray experiment were not validated by the QPCR assay, the expression profiles still yield important information about constitutive expression, the potential alteration of normal developmental expression by the incremental hyperthermia regime, and individual biological variability of expression. The difference between the expression profile of *trypsin-1 precursor* in

the control group of sac fry relative to the experimental group of sac fry suggests that exposure to the hyperthermic challenge may have disrupted constitutive expression of this transcript. The high range in individual biological variability observed at all time points, except the 81 dpf control group, may be a factor contributing to the absence of any significant differences in expression of *chymotrypsin b* within the control or experimental group.

The microarray identified down-regulated digestion-associated transcripts during the hyperthermia regime in both 85 and 89 dpf sac fry, prior to exogenous feeding. *Trypsin-1 precursor* and *chymotrypsin b* were both functionally annotated with 'proteolysis' as a biological process (see Table 3.4, for example). Understanding the constitutive expression of these transcripts may contribute to the understanding of how the disruption of ontogeny of digestive system function by an environmental factor such as temperature may influence fitness. For example, although a dataset of the Atlantic salmon developmental transcriptome has been released (Jantzen *et al.* 2011), *trypsin*-like genes were not present in lists of genes comparable to the developmental stages studied in this experiment. It is possible that, as a factor of the faster growth rate previously observed in Atlantic salmon exposed to hyperthermic conditions (Ytteborg *et al.* 2010), the experimental fish in this study may have up-regulation of the transcript encoding this digestive enzyme at 93 dpf as they develop and approach first-feeding at an escalated rate.

Trypsin and chymotrypsin transcript expression has been previously studied

in larval fish. For example, Murray *et al*. (2004) used RT-PCR to show that *trypsin 1* expression was high in intestine and pyloric caecum of adult winter flounder (*Pleuronectes americanus*), although this transcript was not detectable in larval winter flounder using *in situ* hybridization or RT-PCR. However, in this same study, *trypsin-2* and *trypsin-3* expression was observed using RT-PCR and *trypsin-2* was observed using *in situ* hybridization in larval stages. A previous study on larval development of red porgy (*Pagrus pagrus*) revealed, by *in situ* hybridization, that *trypsinogen* was present from hatching onward and was involved in proteolysis during larval development prior to the onset of exogenous feeding (Darias *et al.* 2007). Similarly, whole mount *in situ* hybridization of Japanese flounder (*Paralichthys olivaceus*) was used to determine that *trypsinogen-1* and *chymotrypsinogen-1* mRNA were expressed at first feeding stage (Srivastava *et al.* 2002; Suzuki *et al.* 2002).

In addition to the aforementioned transcript expression studies, trypsin and chymotrypsin protein expression and/or activity has also been studied in larval fish. For example, the enzyme activity of trypsin and chymotrypsin increased after hatch in Senegalese sole (*Solea senegalensis*) larvae (inert diet fed group; Gamboa-Delgado *et al.* 2011) and increased prior to the onset of exogenous feeding in Japanese eel (*Anguilla japonica*) larvae; interestingly, trypsin and chymotrypsin activity also increased significantly in response to increases in temperature (Murashita *et al.* 2013). Although no induction of these genes was observed in the

control condition within the time frame of the current study, collectively, these prior studies suggest that induction of trypsin and chymotrypsin genes is a conserved component of the development of larval teleost fish as they approach first feeding stage. If the expression of these transcripts is functionally important in Atlantic salmon as they approach first feeding stage, then disruption of ontological expression could influence fitness at this life stage.

Constitutive expression of *ferritin middle subunit* mRNA remained consistent across all time points of the study in the control group. However, *ferritin middle subunit* was microarray identified as up-regulated in 85 dpf sac fry at 11 °C compared to 7 °C (confirmed by QPCR), and was functionally annotated for cellular iron ion homeostasis and ferric iron binding (Table 3.3). Further, in the experimental condition, *ferritin middle subunit* was significantly up-regulated at 11 °C relative to all other time points and at 15 °C relative to lower temperature time points; differences between the constitutive and experimental expression profiles show that the expression of this transcript was influenced by the experimental regime. In teleost fish, the up-regulation of *ferritin* at the transcript level has been elicited in response to bacterial pathogens or antigens (Peatman *et al.* 2007; Neves *et al.* 2009), as well as in rainbow trout in response to a cold temperature shift from 22 °C to 4 °C (Yamashita *et al.* 1996). Ferritin abundance increased in response to short-term heat response (36 °C for four days) in murrel (*Channa striatus*) liver (Mahanty *et al.* 2016). Also, ferritin is known to respond to anoxia stress (Hu *et al.* 2011; Zheng *et*

al. 2010) to increase blood oxygen carrying capacity, and is involved in iron binding and transport, and iron ion homeostasis (Elvitigala *et al.* 2013; Orino *et al.* 2001). In salmon, the up-regulation of *ferritin* transcripts has been associated with growth and the immune response (Xu et al. 2013; Rise et al. 2004). Collectively, these studies show the importance of ferritin during biological processes (e.g. growth, response to pathogens) that are important for survival. The disruption of this transcript (as seen in the current study, which showed the dysregulation of *ferritin* transcript at moderately elevated temperature), could potentially affect these biological processes.

Similar to the profile expression of *ferritin*, there was no significant change in *ubiquitin* transcript expression in the control group (held at 7 °C) for the duration of the study (81 dpf to 93 dpf). *Ubiquitin* was microarray-identified as up-regulated in 89 dpf sac fry at 15 °C compared to 7 °C (confirmed for this time point by QPCR), and *ubiquitin* expression was significantly up-regulated in the experimental condition at 89 dpf (15 °C) compared with all other experimental time points. Although no functional annotation was available for the Atlantic salmon transcript, *ubiquitin* has been known to function during apoptosis and the stress response (Table 3.5; Wen *et al.* 2012; Song *et al.* 2012). Under increased thermal conditions, ubiquitination may be used to mark proteins that have become denatured in response to heat, sending them to the proteasome for proteolysis (Logan and Somero 2011; Song *et al.* 2012). Ubiquitin has been associated with the activity of molecular chaperones, such as the family of sHSPs, which prevent incorrect tertiary folding of proteins denatured by

heat stress (Patterson and Höhfeld 2006), and has also been associated with the acute phase of thermal stress response in mammals (Ahlskog *et al.* 2010; Bayne and Gerwick 2001; Lanneau *et al.* 2010).

3.5.3 Gene expression studies of four paralogous *α-sHSP* transcripts in Atlantic salmon sac fry during chronic incremental hyperthermia

The suite of paralogous *α-sHSP* transcripts whose expression was significantly higher surrounding hatch compared to eye-up stage embryos (Chapter 2) were also of interest in the present study due to their potential functional relevance as known heat stress-responsive genes. The results of the current study described the ontogeny of transcript expression of these paralogues prior to exogenous feeding in late sac fry Atlantic salmon. Although no significant induction of constitutive expression of the *α-sHSPs* was observed, the significant up-regulation of *α-sHSPs 1, 3,* and *4* in the experimental condition suggests that these transcripts are responsive to increased temperature. In this chapter, *α-sHSP 1* and *2,* a pair of paralogues that were closely related, had more similar transcript expression profiles in response to the thermal challenge while *α-sHSP 3* and *4* (which were also closely related) had dissimilar expression profiles. The divergence in expression profiles observed among these highly similar paralogues suggests some sub-functionalization of these genes, as the presence of multiple paralogues would theoretically reduce evolutionary constraint on gene family members.

The various members of the sHSP superfamily of genes have been extensively studied, as has been discussed throughout this thesis. In fish, previous studies have associated these genes with ancestral functions, have studied the duplication and radiation of these gene families, and have identified other stress-related responses and functions (reviewed in Heikkila *et al.* 2017). Members of the sHSP superfamily are heat stress-responsive, acting as chaperones to prevent incorrect tertiary folding of thermally denatured proteins (Heikkila *et al.* 1986; Patterson and Höhfeld 2006; Skjærven *et al.* 2011). For example, *hspb11* (alias *HSP30*) was up-regulated in response to 1 h of 37 °C heat shock in 24 and 48 hpf zebrafish embryos (Marvin *et al.* 2008), and accumulated in response to heat shock as early as 12 hpf (Elicker and Hutson 2007). sHSP26 was induced by heat shock in Pacific abalone (*Haliotis discus hannai*), and a study by Norris *et al.* (1997) determined that hsp27 and hsp30 were responsive to heat shock in a desert minnow, the clearfin livebearer *(Poeciliopsis lucida*); this study also provided evidence of gene duplication events leading to the diversification of these two gene families, and suggested that the hsp30 family of genes underwent more rapid divergence than the hsp27 family in this species due to lower evolutionary constraint. Additionally, *hspb11* was up-regulated in channel catfish gill following infection with the bacterial pathogen *Flavobacterium columnare* (Xie *et al.* 2015). The current study contributes to the literature on *hsp30*-like genes and presented an opportunity to examine the evolution of a well-known family of stress-responsive genes in the pseudotetraploid

Atlantic salmon, where reduced evolutionary constraint among duplicated genes may lead towards rediploidization, sub-functionalization, or neofunctionalization in closely related paralogues.

3.6 CONCLUSIONS

In conclusion, this research identified transcripts and potential biomarkers that were dysregulated in response to incremental hyperthermia. The expression profiles of *trypsin-1 precursor* and *chymotrypsin b* transcripts provided information on the ontogeny of digestive pathways potentially influenced by temperature prior to first feed. The up-regulated expression profile of *ubiquitin* transcript indicates a potential response to the stress of the experimental challenge. The up-regulated expression of *ferritin middle subunit* indicates that the elevated temperature may have altered iron ion homeostasis. Functional conservation and divergence among paralogous genes was observed in the expression profiles of paralogous *α-sHSP* transcripts. The current study contributes to the available data on defence-relevant gene expression in Atlantic salmon sac fry, and identifies candidate molecular biomarkers (with the development of associated QPCR assays) that may be useful in future studies on the impact of elevated temperature on early life stage salmon.

3.7 REFERENCES

Adzhubei, A.A., Vlasova, A.V., Hagen-Larsen, H., Ruden, T.A., Laerdahl, J.K., Høyheim, B. (2007). Annotated expressed sequence tags (ESTs) from pre-smolt Atlantic salmon (*Salmo salar*) in a searchable data resource. *BMC Genomics* 8:209.

Ahlskog, J.K., Björk, J.K., Elsing, A.N., Aspelin, C., Kallio, M., Roos-Mattjus, P., Sistonen, L. (2010). Anaphase-promoting complex/cyclosome participates in the acute response to protein-damaging stress. *Molecular and Cellular Biology* 30(24):5608-5620.

Allendorf, F.W., Thorgaard, G.H. (1984). Tetraploidy and the evolution of salmonid fishes. In: Evolutionary genetics of fishes (Ed. B.J. Turner), Plenum Press, New York, 1-53.

Aluru, N., Vijayan, M.M. (2009). Stress transcriptomics in fish: a role for genomic cortisol signalling. *General and Comparative Endocrinology* 164(2-3):142-150.

Bayne, C.J., Gerwick, L. (2001). The acute phase response and innate immunity of fish. *Developmental and Comparative Immunology* 25:725-743.

Bergheim, A., Fivelstad, S. (2014). Atlantic salmon (*Salmo salar* L.) in aquaculture: metabolic rate and water flow requirements. In: Salmon: Biology, Ecological

Impacts and Economical Importance (Eds. Woo, P.T.K., Noakes, D.J.), Nova Science Publishers, Inc.

Bogner, J., Pipatti, R., Hashimoto, S., Diaz, C., Mareckova, K., Diaz, L., *et al.* (2008). Mitigation of global greenhouse gas emissions from waste: conclusions and strategies from the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report. Working Group III (Mitigation). *Waste Management and Research*, 1:11-32.

Booman, M., Borza, T., Feng, C.Y., Hori, T.S., Higgins, B., Culf, A., *et al.* (2011). Development and experimental validation of a 20K Atlantic cod (*Gadus morhua*) oligonucleotide microarray based on a collection of over 150,000 ESTs. *Marine Biotechnology (NY)* 13(4):733-750.

Danzmann, R.G., Cairney, M., Davidson, W.S., Ferguson, M.M., Gharbi, K., Guyomard, R., *et al.* (2006). A comparative analysis of the rainbow trout genome with 2 other species of fish (Arctic charr and Atlantic salmon) within the tetraploid derivative Salmonidae family (subfamily: Salmoninae). *Genome* 48:1037-1051.

Darias, M.J., Murray, H.M., Gallant J.W., Douglas, S.E., Yufera, M., Martinez-Rodriguez, G. (2007). The spatiotemporal expression pattern of trypsinogen and and bile salt-activated lipase during the larval development of red porgy (*Pagrus pagrus*, Pisces, Sparidae). *Marine Biology* 152(1):109-118.

Davidson, W.S., Koop, B.F., Jones, S.J.M., Iturra, P., Vidal, R., Maass, A., *et al.* (2010). Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biology* 11:403-409.

DeCola, J.N. (1970). Water quality requirements for Atlantic salmon, USDI. Federal Water Quality Administration, Boston, Massachusetts.

Elicker, K.S., Hutson, L.D. (2007). Genome-wide analysis and expression profiling of the small heat shock proteins in zebrafish. *Gene* 403(1-2):60-69.

Elliott, J.M. (1991). Tolerance and resistance to thermal stress in juvenile Atlantic salmon, *Salmo salar*. *Freshwater Biology* 25:61-70.

Elvitigala, D.A., Premachandra, H.K., Whang, I., Oh, M.J., Jung, S.J., Park, C.J., Lee, J. (2013). A teleostean counterpart of ferritin M subunit from rock bream (*Oplegnathus fasciatus*): An active constituent in iron chelation and DNA protection against oxidative damage, with a modulated expression upon pathogen stress. *Fish and Shellfish Immunology.*

Emerson, K., Russo, R.C., Lund, R.E., Thurston, R.V. (1975). Aqueous ammonia equilibrium calculations: effect of pH and temperature. *Journal of the Fisheries Research Board of Canada* 32(12):2379-2383.

Finn, R.N. (2007). The physiology and toxicology of salmonid eggs and larvae in relation to water quality criteria. *Aquatic Toxicology* 8(4): 337-354.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative

mutations. *Genetics* 151(4):1531-1545.

Gamboa-Delgado, J., Le Vay, L., Fernández-Díaz, C., Cañavate, P., Ponce, M., Zerolo, R., Manchado, M. (2011). Effects of different diets on proteolytic enzyme activity, trypsinogen gene expression and dietary carbon assimilation in Senegalese sole (*Solea senegalensis*) larvae. *Comparative Biochemistry and Physiology, Part B: Biochemistry and Molecular Biology* 158:251-258.

Hagen-Larsen, H., Laerdahl, J.K., Panitz, F., Adzhubei, A., Hoyheim, B. (2005). An EST-based approach for identifying genes expressed in the intestine and gills of pre-smolt Atlantic salmon (*Salmo salar*). *BMC Genomics* 6:17.

Hall, J.R., Clow, K.A., Rise, M.L., Driedzic, W.R. (2011). Identification and validation of differentially expressed transcripts in a hepatocyte model of cold-induced glycerol production in rainbow smelt (*Osmerus mordax*). *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology* 301(4):R995-R1010.

Heikkila, J.J. (2017). The expression and function of *hsp30*-like small heat shock protein genes in amphibians, birds, fish, and reptiles. *Comparative Biochemistry and Physiology, Part A: Molecular and Integrative Physiology* 203:179-192.

Heikkila, J.J., Browder, L.W., Gedamu, L., Nickells, R.W., Schultz, G.A. (1986). Heat-shock gene expression in animal embryonic systems. *Canadian Journal of Genetics and Cytology* 28(6):1093-105.

Hori, T.S., Gamperl, A.K., Afonso, L.O., Johnson, S.C., Hubert, S., Kimball, J., *et al.* (2010). Heat-shock responsive genes identified and validated in Atlantic cod (Gadus morhua) liver, head kidney and skeletal muscle using genomic techniques. *BMC Genomics* 11:72.

Hori, T.S., Gamperl, A.K., Booman, M., Nash, G.W., Rise, M.L. (2012). A moderate increase in ambient temperature modulates the Atlantic cod (Gadus morhua) spleen transcriptome response to intraperitoneal viral mimic injection. *BMC Genomics* 13(1):431.

Hu, Y.H., Zheng, W.J., Sun, L. (2011). Identification and molecular analysis of a ferritin subunit from red drum (*Sciaenops ocellatus*). *Fish and Shellfish Immunology* 28(4):678-686.

Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N., Mauceli, E., *et al.* (2004). Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature 431: 946-957.

Jantzen, S.J., Sanderson, D.S., von Schalburg, K.R., Yasuike, Marass, F., Koop, B.F. (2011). A 44K microarray dataset of the changing transcriptome in developing Atlantic salmon (*Salmo salar* L.) *BioMed Central Research Methods* 4:88.

Jonsson, B., Jonsson, N. (2009). A review of the likely effects of climate change on anadromous Atlantic salmon *Salmo salar* and brown trout *Salmo trutta*, with particular reference to water temperature and flow. *Journal of Fish Biology*

75:2381-2447.

Koop, B.F., von Schalburg, K.R., Leong, J., Walker, N., Lieph, R., Cooper, G.A., *et al.* (2008). A salmonid EST genomic study: genes, duplications, phylogeny and microarrays. *BMC Genomics* 9: 545.

Lanneau, D., Wettstein, G., Bonniaud, P. Garrido, C. (2010). Heat shock proteins: cell protection through protein triage. *The Scientific World Journal* 10:1543-1552.

Li, C., Ortí, G., Zhang, G., Lu, G. (2007). A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evolutionary Biology* 7: 44.

Lien, S., Koop, B.F., Sandve, S.R., Miller, J.R., Kent, M.P., Nome, T., *et al.* (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature* doi: 10.1038/nature17164

Livak, K.J., Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25(4):402-408.

Logan, C.A., Somero, G.N. (2011). Effects of thermal acclimation on transcriptional responses to acute heat stress in the eurythermal fish *Gillichtys mirabilis* (Cooper). *American Journal of Physiology – Regulatory, Integrative and Comparative Physiology* 300:1373-1383.

Long, Y., Li. L., Li. Q., He, X., Cui, Z. (2012). Transcriptomic characterization of temperature stress responses in larval zebrafish. *PLoS ONE* 7(5).

Mahanty, A., Purohit, G.K., Banerjee, S., Karunakaran, D., Mohanty, S., Mohanty, B.P. (2016). Proteomic changes in the liver of *Channa striatus* in response to high temperature stress. *Electrophoresis* 37(12):1704-1717.

Marvin, M., O'Rourke, D., Kurihara, T., Juliano, C.E., Harrison, K.L., Hutson, L.D. (2008). Developmental expression patterns of the zebrafish small heat shock proteins. *Developmental Dynamics* 237(2):454-63.

Mirkovic, T., Rombough, P. (1998). The effect of body mass and temperature on the heart rate, stroke volume, and cardiac output of larvae of the rainbow trout, *Oncorhynchus mykiss*. *Physiological Zoology* 71(2): 191-197.

Murashita, K., Furuita, H., Matsunari, H., Yamamoto, T., Awaji, M., Nomura, K., Nagao, J., Tanaka, H. (2013). Partial characterization and ontogenic development of pancreatic digestive enzymes in Japanese eel *Anguilla japonica* larvae. *Fish Physiology and Biochemistry* 39:895-905.

Murray, H.M., Pérez-Casanova, J.C., Gallant, J.W., Johnson, S.C., Douglas, S.E. (2004). Trypsinogen expression during the development of the exocrine pancreas in winter flounder (*Pleuronectes americanus*). *Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology.* 138(1):53-59.

Neves, J.V., Wilson, J.M., Rodrigues, P.N.S. (2009). Transferrin and ferritin response to bacterial infection: the role of liver and brain in fish. *Developmental and Comparative Immunology* 33:848-857.

Norris, C.E., Brown, M.A., Hickey, E., Weber, L.A., Hightower, L.E. (1997). Low-molecular-weight heat shock proteins in a desert fish (*Poeciliopsis lucida*): homologs of human Hsp27 and Xenopus Hsp30. *Molecular Biology and Evolution.* 14(10):1050-1061.

Ohno, S. (1970). Evolution by gene duplication. Springer-Verlang, Heidelberg, Germany.

Orino, K., Lehman, L., Tsuji, Y., Ayaki, H., Torti, S.V., Torti, F.M. (2001). Ferritin and the response to oxidative stress. *Biochemical Journal* 357:241-247.

Park, E.M., Kim, Y.O., Nam, B.H., Kong, H.J., Kim, W.J., Lee, S.J., Jee, Y.J., Kong, I.S., Choi, T.J. (2008). Cloning and expression analysis of a small HSP26 gene of Pacific abalone (*Haliotis discus hannai*). *Journal of Environmental Biology* 29(4):577-580.

Patterson, C., Höhfeld, J. (2006). Molecular chaperones and the ubiquitin-proteasome system. In: Protein Degradation: The Ubiquitin-Proteasome System, Volume 2 (Eds. R. J. Mayer, A. J. Ciechanover and M. Rechsteiner), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

Peatman, E., Baoprasertkul, P., Terhune, J., Xu, P., Nandi, S., Kucuktas, H., *et al.*

(2007). *Developmental and Comparative Immunology* 31:1183-1196.

Pérez-Casanova, J.C., Rise, M.L., Dixon, B., Afonso, L.O., Hall, J.R., Johnson, S.C., Gamperl, A.K. (2008). The immune and stress responses of Atlantic cod to long-term increases in water temperature. *Fish and Shellfish Immunology* 24: 600-609.

Peterson, R.H., Spinney, H.C.E., Sreedharan, A. (1977). Development of Atlantic salmon (*Salmo salar*) eggs and alevins under varied temperature regimes. *Journal of the Fisheries Research Board of Canada* 34:31-43.

Ravi, V., Venkatesh, B. (2008). Rapidly evolving fish genomes and teleost diversity. *Current Opinion in Genetics & Development* 18: 544-550.

Rise, M.L., Hall, J., Rise, M., Hori, T., Gamperl, A.K., Kimball, J., *et al.* (2008). Functional genomic analysis of the response of Atlantic cod (*Gadus morhua*) spleen to the viral mimic polyriboinosinic polyribocytidylic acid (pIC). *Developmental ad Comparative Immunology* 32(8): 916-931.

Rise, M.L., Jones, S.R.M., Brown, G.D., von Schalburg, K.R., Davidson, W.S., Koop, B.F. (2004a). Microarray analyses identify molecular biomarkers of Atlantic salmon macrophage and hematopoietic kidney response to *Piscirickettsia salmonis* infection. *Physiological Genomics* 20:21-35.

Rise, M.L., von Schalburg, K., Brown, G.D., Devlin, R.H., Mawer, M.A., Kuipers, N., *et al.* (2004b). Development and application of a salmonid EST database

and cDNA microarray: data mining and interspecific hybridization characteristics. *Genome Research* 14:478-490.

Rozen, S., Skaletsky, H.J. (2000). Primer3 on the WWW for general users and for biologist programmers. Bioinformatics Methods and Protocols: Methods in Molecular Biology (Eds. Krawetz, S., Misener, S.), Humana Press, New Jersey, 365-386.

Scott, W.B., Scott, M.G. (1988). Atlantic Fishes of Canada. *Canadian Bulletin of Fisheries and Aquatic Sciences* 219:129-134.

Skjærven, K.H., Olsvik, P.A., Finn, R.N., Holen, E., Hamre, K. (2011). Ontogenic expression of maternal and zygotic genes in Atlantic cod embryos under ambient and thermally stressed conditions. *Comparative Biochemistry and Physiology, Part A: Molecular and Integrative Physiology* 159:196-205.

Song, Y., Salbu, B., Heier, L.S., Teien, H.S., Lind, O.C., Oughton, D., *et al.* (2012). Early stress response in Atlantic salmon (*Salmo salar*) exposed to environmentally relevant concentrations of uranium. *Aquatic Toxicology* 112-113:62-71.

Srivastava, A.S., Kurokawa, T., Suzuki, T. (2002). mRna expression of pancreatic enzyme precursors and estimation of protein digestibility in first feeding larvae of the Japanese flounder, *Paralichthys olivaceus*. *Comparative Biochemistry and Physiology, Part A: Molecular and Integrative Physiology* 132:629-635.

Suzuki, T., Srivastava, A.S., Kurokawa, T. (2002). cDNA cloning and phylogenetic analysis of pancreatic serine proteases from Japanese flounder, *Paralichthys olivaceus*. *Comparative Biochemistry and Physiology, Part B: Biochemistry and Molecular Biology* 131(1):63-70.

Tort, L. (2011). Stress and immune modulation in fish. *Developmental and Comparative Immunology* 35(12):1366-1375.

von Schalburg, K.R., Rise, M.L., Cooper, G.A., Brown, G.D., Gibbs, A.R., Nelson, C.C., *et al.* (2005). Fish and chips: various methodologies demonstrate utility of a 16,006-gene salmonid microarray. *BMC Genomics* 6:126.

Wei, J., Carroll, R.J., Harden, K.K., and Wu, G. (2012). Comparisons of treatment means when factors do not interact in two-factorial studies. *Amino Acids* 42(5):2031-2035.

Wolfe, K.H. (2001). Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics* 2:333-341.

Xu, Q., Feng, C.Y., Hori, T.S., Plouffe, D.A., Buchanan, J.T., Rise, M.L. (2013). Family-specific differences in growth rate and hepatic gene expression in juvenile triploid growth hormone (GH) transgenic Atlantic salmon (*Salmo salar*). *Comparative Biochemistry and Physiology, Part D: Genomics and Proteomics* 8(4):317-33.

Yamashita, M., Ojima, N., Sakamoto, T. (1996). Molecular cloning and

cold-inducible gene expression of ferritin H subunit isoforms in rainbow trout cells. *Journal of Biological Chemistry* 271(43):26908-26913.

Ytteborg, E., Baeverfjord, G., Torgersen, J., Hjelde, K., Takle, H. (2010). Molecular pathology of vertebral deformities in hyperthermic Atlantic salmon (*Salmo salar*). *BMC Physiology* 10:12.

Zheng, W., Hu, Y., Sun. L. (2010). Identification and analysis of a *Scophthalmus maximus* ferritin that is regulated at transcription level by oxidative stress and bacterial infection. *Comparative Biochemistry and Physiology, Part B: Biochemistry and Molecular Biology* 156:222-228.

3.8 WEB REFERENCES

3.1 http://blast.ncbi.nlm.nih.gov/Blast.cgi; BLAST: Basic Local Alignment Search Tool.

3.2 http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi; NCBI Taxonomy Browser.

3.3 http://frodo.wi.mit.edu/primer3/; Primer3 V.4.0.

3.4 http://www.uniprot.org/; UniProt Knowledgebase.

3.5 http://edis.ifas.ufl.edu/fa031; Water Quality: Ammonia in Aquatic Systems

3.6 http://ceqg-rcqe.ccme.ca/download/en/141; Water Quality: Canadian Water Quality Guidelines for the Protection of Aquatic Species.

3.7 https://www.dep.state.fl.us/labs/docs/unnh3disc.docx; Water Quality: Explanation of the Discrepancy between the Florida Department of Environmental Protection's Un-ionized Ammonia Calculator (Version 2.1) and the EPA Text Table 3 Values in EPA 440/5-88-004

4. FUTURE RESEARCH

As genomic technology advances and genomic resources for Atlantic salmon continue to improve (e.g., availability of a complete genome sequence; Lien *et al.* 2016), our ability to study the genetic basis of salmonid development and early life stage salmon responses to environmental stressors will improve. Although the 32K cGRASP salmonid cDNA microarray represented the best coverage of the Atlantic salmon transcriptome at the time that this research was conducted, RNA-sequencing (RNA-seq) will likely be the transcriptomic method of choice for future studies. The availability of a complete genome sequence for Atlantic salmon facilitates RNA-seq, as well as other molecular studies in the species (e.g., complete characterization of genes) (Lien *et al.* 2016).

Limitations of these experiments should be considered and implicated in the experimental design of future research. The developmental expression study in Chapter 2 of this thesis would be strengthened by repeating the QPCR analyses with a separate batch of salmon embryos (e.g., arising from different parents) to determine if the observed expression results are repeatable. Confidence in the interpretation of the results of the thermal challenge study in Chapter 3 of this thesis would be strengthened by more a similar design between the experimental and control conditions (e.g., by running each condition in side-by-side diurnal incubators). Future research in this area could also benefit from cortisol assays, to determine if a stress response is mounted in the salmon sac fry exposed to incremental

hyperthermia. For future studies, a greater number of biological replicates at each sampling point could increase statistical power of QPCR results. Research designed to incorporate multiple species from various branches of the evolutionary tree under a common experimental paradigm could also aid in the interpretation of results involving paralogous suites of transcripts, the notion of pre-adaptation to hatching stress, and adaptability to elevated temperatures during early life stages.

Further studies of Atlantic salmon early development could lead to a more complete understanding of the transcriptome dynamics during embryogenesis, surrounding the hatch event, and during sac fry development, and could potentially shed light on the means by which the embryos prepare to emerge from the egg shell and cope with the challenges faced in their new environment. Future research could involve how early response to environmental stressors and immune challenges may alter the development and physiology (function) of the Atlantic salmon digestive and immune systems. Aside from the inherent evolutionary insights these types of studies could provide, defence-relevant biomarkers for Atlantic salmon embryo and sac fry could also aid in the development of aquaculture rearing techniques. Understanding the innate immune and defence strategies of this species, and the ability to cope with and adapt to various environmental challenges, may also contribute to research on the development and survival of Atlantic salmon embryos in the face of accelerated climate change and other anthropogenic influences.

5. APPENDIX

Table A.1 Summary of statistical analysis (one-way ANOVA, followed by Tukey post-hoc tests) of the expression of eight genes of interest (examined by QPCR) in Atlantic salmon during early development from eye-up stage (45 dpf) to post-hatch stage (61 dpf).

Gene of Interest		**cryGM2-like 1**	**cryGM2-like 2**	**cryGM4-like**	**cryGm3-like**	**α-sHSP 1**	**α-sHSP 2**	**α-sHSP 3**	**α-sHSP 4**
Test/Comparison									
p-value ANOVA		0.033*	0.184	0.011*	0.008*	0.001*	0.001*	<0.001*	<0.001*
p-value Tukey	**45 dpf vs. 49 dpf**	0.991	0.968	0.997	0.092	1.00	1.00	0.998	0.973
	45 dpf vs. 53 dpf	0.512	0.485	0.735	0.412	0.686	0.217	0.390	0.636
	45 dpf vs. 57 dpf	0.114	0.208	0.385	0.053	0.003*	0.002*	<0.001*	<0.001*
	45 dpf vs. 61 dpf	0.054	0.472	0.016*	0.023*	0.043*	0.081	<0.001*	<0.001*
	49 dpf vs. 53 dpf.	0.757	0.765	0.882	0.622	0.597	0.141	0.586	0.916
	49 dpf vs. 57 dpf	0.263	0.376	0.534	0.094	0.002*	0.001*	<0.001*	0.001
	49 dpf vs. 61 dpf	0.137	0.750	0.002*	0.040	0.049*	0.056	<0.001*	<0.001*
	53 dpf vs. 57 dpf	0.968	0.970	0.962	0.758	0.130	0.049*	0.001*	0.004*
	53 dpf vs. 57 dpf	0.857	1.00	0.120	0.465	0.819	0.974	0.001*	<0.001*
	57 dpf vs. 61 dpf	0.992	0.974	0.365	0.978	0.240	0.313	1.00	0.583

* $p < 0.05$

Table A.2 *Oncorhynchus mykiss α-sHSP* paralogous contiguous sequences (contigs) and numbers of contributing ESTs (with associated cDNA libraries) identified and assembled in the cGRASP EST database.

***O. mykiss α-sHSP* Gene (contig #)[1]**	**Contig Length (bp)**	**cGRASP cDNA Library Identifier[2]**	**Tissue Used to Generate cDNA Library**	**No. of Clones[3]**
***O.mykiss* α-sHSP 1 (cons26785)**	945	omyk_rbna	whole juvenile	2
		omyk_rbnb	whole juvenile	3
***O.mykiss* α-sHSP 2 (cons30611)**	1140	omyk_rbna	whole juvenile	1
		omyk_rbnb	whole juvenile	2
***O.mykiss* α-sHSP 3 (cons111289)**	878	omyk_rbnb	whole juvenile	6
***O.mykiss* α-sHSP 4 (cons21483)**	972	omyk_rbha	whole juvenile	1
		omyk_rbna	whole juvenile	1
		omyk_rbnb	whole juvenile	2

[1]Contig numbers (cons#) were automatically assigned by the cGRASP EST database (Web Reference 2.3).

[2]The identifiers (names) of the cDNA libraries in the cGRASP EST database (Web Reference 2.3) were provided in Koop *et al*. (2008).

[3]Some clones were sequenced more than once (forward and reverse), so that the number of clones is not necessarily equal to the number of EST sequences.

Table A.3 Two-way ANOVA of percent dissolved O_2 saturation over the experimental period (*Time*) and between the control and experimental treatments (*Condition*).

Source	Type III Sum of Squares	df	Mean Square	F	p-value
Model	901718.667 [1]	30	30057.289	23320.310	<0.0001*
Condition	74.711	1	74.711	57.966	<0.0001*
Time	21.622	14	1.544	1.198	0.301
Condition* x *Time	21.622	14	1.544	1.198	0.301
Error	77.333	60	1.289		
Total	901796.000	90			

[1]R Squared = 1.000 (Adjusted R Squared = 1.000)
* P < 0.05

Table A.4 Mann-Whitney U Test of the effects of *Condition* (control or experimental) on percent dissolved oxygen saturation.

Condition	N	Mean Rank	Sum of Ranks	U	p-value
Control	45	62.5	2812.5	247.500	<0.0001*
Experimental	45	28.5	1282.5		

* P < 0.05

Table A.5 Two-way ANOVA on unionized ammonia levels (NH_3) over the experimental period (*Time*) and between the control and experimental treatments (*Condition*).

Source	Type III Sum of Squares	df	Mean Square	F	p-value
Model	1.309E-6 [1]	30	4.363E-08	1237.885	<0.0001*
Day	1.065E-07	14	7.608E-09	215.883	<0.0001*
Treatment	1.510E-07	1	1.510E-07	4283.306	<0.0001*
Day* x *Treatment	1.047E-07	14	7.479E-09	212.224	<0.0001*
Error	2.115E-09	60	3.524E-11		
Total	1.311E-06	90			

[1]R Squared = .998 (Adjusted R Squared = .998)
* P < 0.05

Table A.6 One-way ANOVA of the combined effects of *Time* and *Condition* on NH_3.

Sum of Squares		df	Mean Square	F	p-value
Between Groups	3.622E-07	29	1.249E-08	354.372	<0.0001*
Within Groups	2.115E-09	60	3.524E-11		
Total	3.643E-07	89			

* P < 0.05

Table A.7 Levene's test of homogeneity of variances for NH_3 data.

Levene Statistic	df1	df2	Sig.
1.641	29	60	0.053

Table A.8 Two-way ANOVA on mortality accumulated over the experimental period (*Time*) and the differences in cumulative mortality between the control and experimental treatment (*Condition*).

Source	Type III Sum of Squares	df	Mean Square	F	p-value
Model	49105.000 [1]	30	1636.833	288.853	<0.0001*
Condition	6812.100	1	6812.100	1202.135	<0.0001*
Time	16117.622	14	1151.259	203.163	<0.0001*
Condition * Time	9008.067	14	643.433	113.547	<0.0001*
Error	340.000	60	5.667		
Total	49445.000	90			

[1]R Squared = .993 (Adjusted R Squared = .990)
* P < 0.05

Table A.9 One-way ANOVA of the combined effects of *Time* and *Condition* on cumulative mortality.

Sum of Squares		df	Mean Square	F	p-value
Between Groups	31937.78889	29	1101.303	194.348	<0.0001*
Within Groups	340.000	60	5.667		
Total	32277.78889	89			

* P < 0.05

Table A.10 Levene's test of homogeneity of variances for cumulative mortality data.

Levene Statistic	df1	df2	p-value
3.890	29	60	<0.0001*

* P < 0.05

Table A.611 Summary of statistical analysis (two-way ANOVA, followed by Student-Newman-Keuls post-hoc tests) of the expression of eight genes of interest (examined by QPCR) in Atlantic salmon late sac fry in response to incremental hyperthermia.

Gene of Interest		**Trypsin-1 precursor**	**Ferritin, middle subunit**	**Chymo-trypsin B**	**Ubiquitin**	**α-sHSP 1**	**α-sHSP 2**	**α-sHSP 3**	**α-sHSP 4**
Test/Comparison									
p-value ANOVA	**Control vs. Experi-mental**	0.076	<0.001*	0.339	0.002*	0.103	0.409	<0.001*	0.261
p-value Student-Newman-Keuls	**81 dpf 7°C vs. 7°C**	0.066	0.247	0.012*	0.894	0.997	0.993	0.853	0.936
	85 dpf 7°C vs. 11°C	0.534	<0.001*	0.932	0.184	0.727	0.781	0.021*	0.898
	89 dpf 7°C vs. 15°C	0.863	0.063	0.568	<0.001*	0.593	0.837	0.002*	0.951
	93 dpf 7°C vs. 19°C	0.015*	0.181	0.746	0.322	0.021*	0.115	0.114	0.048*
p-value Student-Newman-Keuls (Within Control)	**81 dpf vs. 85 dpf**	0.069	0.849	0.080	0.859	0.881	0.631	0.379	0.986
	81 dpf vs. 89 dpf	0.104	0.735	0.182	0.954	0.463	0.527	0.324	0.981
	81 dpf vs. 93 dpf	0.162	0.805	0.091	0.777	0.765	0.790	0.304	1.00
	85 dpf vs. 89 dpf	0.486	0.842	0.456	0.974	0.528	0.785	0.723	0.999
	85 dpf vs. 93 dpf	0.336	0.767	0.638	0.904	0.807	0.940	0.806	1.00
	89 dpf vs. 93 dpf.	0.829	0.752	0.892	0.966	0.483	0.501	0.750	0.996
p-value Student-Newman-Keuls (Within Experi-mental)	**81 dpf vs. 85 dpf**	0.825	<0.001*	0.810	0.199	0.260	0.318	0.006*	0.999
	81 dpf vs. 89 dpf	0.629	0.029*	0.371	<0.001*	0.099	0.387	0.001*	0.995
	81 dpf vs. 93 dpf	0.069	0.073	0.260	0.437	0.041*	0.072	0.007*	0.240
	85 dpf vs. 89 dpf	0.751	<0.001*	0.393	0.001*	0.355	0.752	0.455	0.971
	85 dpf vs. 93 dpf	0.038*	<0.001*	0.374	0.915	0.200	0.298	0.705	0.051
	89 dpf vs. 93 dpf.	0.043*	0.564	0.926	0.003*	0.338	0.241	0.509	0.113

* $p < 0.05$